Your Loss Is My Gain: A Recruitment Experiment With Framed Incentives*

Jonathan de Quidt†

April 2, 2014

Latest version available here

Abstract

Empirically, labor contracts that financially penalize failure induce higher effort provision than economically identical contracts presented as paying a bonus for success, an effect attributed to loss aversion. This is puzzling, as penalties are infrequently used in practice. The most obvious explanation is selection: loss averse agents are unwilling to accept such contracts. I formalize this intuition, then run an experiment to test it. Surprisingly, I find that workers were 25 percent *more* likely to accept penalty contracts, with no evidence of adverse or advantageous selection. Consistent with the existing literature, penalty contracts also increased performance on the job by 0.2 standard deviations. I outline extensions to the basic theory that are consistent with the main results, but argue that more research is needed on the long-term effects of penalty contracts if we want to understand why firms seem unwilling to use them.

Keywords: loss aversion; reference points; framing; selection; Mechanical Turk

JEL Classification: D03, J41, D86

*I am grateful to STICERD for financial support, and to many people for helpful discussions, especially Oriana Bandiera, Tim Besley, Gharad Bryan, Tom Cunningham, Greg Fischer, Maitreesh Ghatak, Matthew Levy, George Loewenstein and Torsten Persson. Gabriele Paolacci, Puja Singhal and Kelly Zhang were particularly helpful in setting up the experiment. A 2013 version of this paper was in limited circulation under the title "Recruiting Workers Under Framed Incentives: An Online Labor Market Experiment."

†IIES and London School of Economics. Email address: jonathan.dequidt@iies.su.se.

Consider two otherwise identical job offers, of which the first pays a base wage of $100, plus a bonus of $100 if a performance target is reached, while the second pays a base wage of $200, minus a penalty of $100 if the target is not reached. Rational agents will behave identically under either of these two contracts. However, a large body of empirical evidence suggests that behavior does respond to framing manipulations such as this. In particular several lab and field studies find that workers exert higher effort under the penalty framed contract than the bonus framed one. The leading explanation for these findings is reference dependence and loss aversion (Kahneman and Tversky (1979)), where the frame influences the reference point. The low base pay of the bonus frame sets the agent's reference point low, so bonuses are perceived as gains, while the high base pay of the penalty frame sets her reference point high, so penalties are perceived as losses. Since losses loom larger than gains, penalties are more motivating than bonuses.

If penalties are more motivating than bonuses, why are they not more widely used by firms?[1] The most obvious explanation is that while penalties are effective motivators of existing workers, they are unlikely to be accepted at the recruitment stage. I formalize this intuition in a simple model, showing that a forward-looking loss averse agent who is subject to framing effects will be less willing to accept a penalty contract than an equivalent bonus contract. This is because penalties increase her reference point, reducing her utility in all states of the world. Under the bonus contract she feels elated when successful and not too disappointed when unsuccessful, while under the penalty contract she feels only contented when successful and very disappointed when unsuccessful.[2] This effect is sufficiently strong that the principal tends to prefer bonuses.

The main contribution of this paper is an online real-effort randomized experiment with 1,450 participants, designed to test this prediction. I use a two-stage design that separates selection and incentive effects of penalty framing relative to bonus framing. It consists of a common first stage in which workers gain experience at the task and I measure their types, followed by a second stage a few days later in which workers are offered incentive contracts that were framed either in terms of bonuses or penalties, and decide whether or not to accept. In each stage workers perform a data entry task and are assessed on their accuracy.

The most striking finding is that, in contrast with the theoretical predictions, workers offered a penalty framed contract were 25 percent *more* likely to accept than those offered an equivalent bonus contract. Second, despite the large effect on recruitment, the penalty contract did not lead to adverse or advantageous selection, indeed the empirical distributions of the main payoff-relevant observables are essentially identical between those who accepted the bonus and those who accepted the penalty. Third, consistent with the existing literature, performance on the incentivized task was sig-

[1] See e.g. Baker et al. (1988), Lazear (1995). Although I am not aware of any datasets addressing this issue, a glance through any job vacancy listing reveals many jobs that specify potential bonuses and almost no mention of penalties.

[2] More generally, in almost any model where she chooses her effort provision optimally, a manipulation that increases effort without changing the economic terms of the contract must make her worse off.

nificantly higher under the penalty treatment, around 6 percent higher accuracy on the data entry task (0.2 standard deviations). The coefficient estimate is unchanged when including controls, reaffirming the absence of selection on observables. The effect sizes are large relative to those for standard manipulations of incentive size: increasing the non-contingent pay from \$0.50 to \$2 increased the acceptance rate by 36 percent and performance by 0.2 standard deviations, while the effect on both outcomes of increasing the contingent component of pay from \$1.50 to \$3 was small and statistically insignificant.

While I cannot of course rule out selection on unobservables, the range of controls used gives confidence that the observed effect is an incentive effect and not driven by selection. I *do* observe significant selection on ability into the incentivized second stage of the experiment (low ability participants are less likely to accept the job offer under both bonus and penalty framing), and I do observe that increasing the non-contingent component of pay in the second stage led to adverse selection.

In addition to controlling for selection, the experiment is designed to rule out two key confounds. Workers are experienced in the task and informed about their ability when deciding whether to accept the job, to avoid them inferring, for example, that a penalty-framed job is easier or harder than a bonus-framed job. I check for such inference effects by testing whether workers perceived the task to be more or less difficult under the penalty frame, and find no difference. Second, I vary the phrasing of the job offers to check whether inattention when reading the offer is driving the results.

Since the basic theory cannot explain the relative popularity of the penalty contract I outline two extensions that bring the model in line with the data. One possibility is that that workers like penalty contracts because they enable them to overcome a self-control problem: the worker would like to exert more effort and the penalty motivates her to do so.[3] Alternatively, it could be that workers are simply failing to correctly translate the terms of the contract into outcomes, and are overly attracted by the high "base pay" under the penalty framed job offer.[4] It turns out that differences in actual pay between bonus and penalty contracts, driven by performance differences, are too small to plausibly explain the large difference in acceptance rates. Furthermore, survey evidence suggests that the penalty contracts were perceived as "more generous", which I interpret as supporting the second proposed mechanism.

There are two possible responses to the results. It may indeed be that firms can gain by increasing their use of penalties, including pre-announcing them at the recruitment stage. This is most likely to be the case in contexts similar to the experimental environment, short-term recruitment of workers to perform routine tasks with minimal screening.

The second response is to go back to the original question. The penalty contract recruited more workers, who then exerted greater effort, why then are they not more

[3]Kaur et al. (2013) find that workers in their sample prefer financially dominated incentive schemes that incorporate a form of self-commitment.

[4]For example, the mechanism may be similar to how eBay buyers seem not to decrease their bids one-to-one in response to an increase in shipping costs, as found by Hossain and Morgan (2006).

widely used? The results suggest that selection is not the answer. Perhaps the explanation lies in the fact that while most employment relationships are long-term, the effects of framing manipulations may be short-lived. Over time, workers' reference points are likely to adjust, eroding the performance advantages and perhaps leading workers who were recruited under the penalty frame but would not have accepted the bonus frame to quit.[5]

Existing work on incentive framing focuses on incentive effects, that is, its effect on effort provision among a sample of already-recruited workers or lab subjects.[6] Hossain and List (2012) in the field, and Armantier and Boly (2012), Hannan et al. (2005) and Church et al. (2008) in the lab consistently find higher effort provision under penalty incentives than equivalent bonus incentives.[7] However, Fehr and Gächter (2002) find in a buyer-seller experiment that penalty-framed performance incentives led to more shirking among sellers than equivalent bonus-framed offers.

The paper relates to the literature on behavioral contract design.[8] In particular de Meza and Webb (2007) and Herweg et al. (2010) study incentives for loss-averse agents without framing effects, while Just and Wu (2005) and Hilken et al. (2013) theoretically analyze an incentive framing problem closely related to the one outlined in this paper. Empirical papers studying the effect of loss aversion on effort provision (without framing effects) include Camerer et al. (1997), Farber (2005, 2008), Crawford and Meng (2011), Pope and Schweitzer (2011), Abeler et al. (2011) and Gill and Prowse (2012).

Finally, it fits into the smaller empirical literature on selection effects of employment contracts. For example, Lazear (2000), Eriksson and Villeval (2008) and Dohmen and Falk (2011) find that performance pay tends to select in high-ability types, a result that I also observe in my experiment, while Guiteras and Jack (2014) observe adverse selection.

The remainder of the paper is as follows. Section 1 sets up the basic theoretical framework and derives three testable predictions. Section 2 outlines the experiment design, the experimental platform (Amazon Mechanical Turk), and the data collected. Section 3 describes the main results on acceptance rates, selection and performance. Section 4 presents two tests of possible mechanisms: inference and inattention. Section 5 discusses extensions to the model that bring it closer in line with the data, and suggestive evidence from a follow-up survey. It also discusses areas where further theoretical

[5] Druckman (2001), Hossain and List (2012) and Jayaraman et al. (2014) discuss the issue of short-lived "behavioral" effects.

[6] The one exception of which I am aware is Luft (1994). In her study, lab participants indicated a preference between each of an increasing sequence of fixed payments, and a contingent contract which was either bonus or penalty framed. The mean valuation of the bonus contract was higher in one treatment, and lower in another, although the sample sizes are small so robust statistical inference on this outcome is difficult. Brooks et al. (2013) study only penalty framed incentives, varying the size of the target below which penalties are incurred. They show that setting the target extremely high reduces acceptance of the job offer and performance.

[7] Fryer et al. (2012) test a closely related but stronger manipulation than pure framing on school teachers: in the penalty treatment teachers were paid their bonuses upfront, to be clawed back if student performance fell below target. They find strong positive effects on teacher performance under the penalty relative to the bonus equivalent.

[8] See Kőszegi (2014) for a recent review.

and empirical work would be particularly valuable. Finally, Section 6 concludes. Three appendices contain additional results and experimental details.

1 A simple model

Consider a standard moral hazard problem in which a principal (P) wants to hire an agent (A) to perform a task, the success of which depends on the agent's effort. Effort is non-contractible so P must write a contract that incentivizes effort. However, A's utility is reference-dependent and loss-averse, and the principal can influence the her reference point by altering how the contract is framed. Lastly, there may be limited liability, such that the payment to the agent in any state must exceed some lower bound $\underline{w}$. A chooses an effort level $e \in [0,1]$ which equals the probability that the task is successful. If successful, P earns a payoff v, otherwise he earns 0.

In the absence of any framing effect, this implies that the optimal contract consists of a pair, (w,b), where w is a non-contingent payment, and b is a bonus for success. Additionally, I assume that P can choose a frame, $F \in [0,1]$, that influences A's reference point under the contract. Thus, P offers A a triple, (w,b,F), A accepts or rejects the contract, then exerts effort if she accepted and is paid according to the outcome. P's payoff is simply $\Pi \equiv e(v-b) - w$.

Following Kőszegi and Rabin (2006, 2007) (henceforth, KR), I assume that the agent's utility function is a sum of a standard component, expected consumption utility, and a gain-loss component that evaluates payoffs against a reference point, less the cost of effort. To keep the presentation simple I follow KR in assuming no probability weighting or diminishing sensitivity. I also assume that the reference point is non-stochastic and determined entirely by the frame, that the cost of effort is not reference-dependent. In Appendix A I work out the implications of allowing the reference point to depend on A's expected effort (using a generalization of KR), diminishing sensitivity, and reference dependence in effort, obtaining essentially the same predictions.

I assume that A's reference point r is equal to the "base pay" specified in the contract, which is $w + Fb$. Thus $F = 0$ corresponds to a pure bonus contract (base pay is equal to w) and $F = 1$ to a pure penalty contract (base pay equal to $w + b$). Intermediate values of F correspond to mixed frames, incorporating both a bonus for success and a penalty for failure. Consumption utility and gain-loss utility are equally weighted. For reference point r, her gain-loss utility if she earns x is equal to $x - r$ if $x \geq r$ (a gain), and $\lambda(x-r)$ if $x < r$ (a loss). $\lambda > 1$ implies that she is loss-averse: losses loom larger than equivalent sized gains.

If she exerts effort e, A receives w with probability $1-e$ and $w+b$ with probability e. The cost of effort is quadratic, depending on an ability parameter γ, less an intrinsic motivation term parameterized by $\alpha < \bar{\alpha}$. Her expected utility is:

$$U(e,w,b,F) = \underbrace{w + eb - \left(\frac{e^2}{2\gamma} - \alpha e\right)}_{\text{Consumption \& cost of effort}} + \underbrace{e(w+b-(w+Fb)) + \lambda(1-e)(w-(w+Fb))}_{\text{Gain-loss utility}}$$

$$= w + e[\alpha + b(2 + (\lambda - 1)F)] - \frac{e^2}{2\gamma} - \lambda F b \tag{1}$$

Given a contract (w, b, F), A's optimal effort choice is equal to:[9]

$$e^*(b,F) = \gamma[\alpha + b(2 + (\lambda - 1)F)] \tag{2}$$

so her maximized utility is equal to:

$$U^*(w,b,F) \equiv w + \frac{\gamma[\alpha + b(2 + (\lambda - 1)F)]^2}{2} - \lambda F b \tag{3}$$

Lastly, A accepts a given contract if her participation constraint is satisfied:

$$U^*(w,b,F) - \bar{u} \geq 0 \tag{4}$$

I assume that $\bar{u}$ is fixed but may depend on A's type $(\gamma, \lambda, \alpha)$.

The model yields three key testable predictions:

Prediction 1 *Suppose A is loss averse ($\lambda > 1$). Then, her effort is higher under a penalty framed contract than an economically equivalent bonus framed contract. I.e. $\frac{de^*}{dF} > 0$.*

Prediction 2 *All else equal, penalties have a larger effect on effort for more loss-averse agents (those with higher λ). I.e. $\frac{d^2e^*}{dF\partial\lambda} > 0$.*

Prediction 3 *A is less willing to accept a penalty contract than the equivalent bonus contract. I.e. $\frac{dU^*}{dF} < 0$.*

It is important to note that without imposing further structure on the outside option, the model does not make specific predictions on which types of agents are more likely be selected out by penalties, i.e. to reject a penalty contract but accept the equivalent bonus contract. Formally, without knowing the distribution of $\bar{u}$ conditional on type, we do not know for whom $U^*(w,b,1) < \bar{u} \leq U^*(w,b,0)$.

The implication of Prediction 3 for optimal contracts is as follows:

Proposition 1 *Suppose P wishes to recruit one agent of known type. P prefers bonus framing to penalty framing whenever A's limited liability constraint is not binding.*

The proof of Proposition 1 simply applies the fact that the non-binding limited liability condition implies that A's participation constraint is binding, and hence P is

[9] I assume that $\gamma < \bar{\gamma} \equiv \frac{1}{\bar{\alpha}+v[1+\lambda]}$ which ensures that $e^* < 1$ in equilibrium (since we know that P will always set $b \leq v$).

the residual claimant of any surplus generated by the relationship. Since by Prediction 3 A's utility is decreasing in F, P will prefer bonus frames, since a marginal decrease in F can be offset by a decrease in w. The complement of this proposition is that P will only use some form of penalty framing ($F > 0$) when the participation constraint is not binding, i.e. when the limited liability constraint ($w \geq 0$) is binding.[10]

2 Experimental design

The experiment was conducted with online workers on MTurk. To separate selection and incentive effects I use a two-stage design similar to Dohmen and Falk (2011). In the first stage, workers are surveyed and perform a practice task under flat incentives, enabling me to measure their types, and giving them experience at the task.[11] In the second stage they are then offered an opportunity to perform the same task under randomly varied performance-related incentives (which depend upon their accuracy), which they can accept or reject. Their outside option is simply determined by their value or leisure or alternative jobs available to them. Selection effects can then be examined by comparing the types that accept different offers, while incentive effects are estimated by comparing behavior conditional on acceptance and type.[12]

A key concern was that agents might interpret the contract offered as informative about the task, or about characteristics of the principal that are relevant to their payoff. For instance, in my context it seems intuitive that agents might perceive a penalty as designed to punish failure at an easy task (or where shirking is easy to detect).[13] Participants might also believe that the choice of bonus or penalty reflects the principal's transaction costs, whereby a bonus (penalty) is chosen when bonuses will be paid (penalties deducted) infrequently, to minimize the number of transactions conducted.

To address these concerns, workers were sent an email informing them of the percentage of strings that they typed correctly in the first stage of the experiment. This accuracy rate measure maps directly into their probability of receiving the bonus (avoiding the penalty) in the second stage, giving them a good signal of their ability or the task difficulty. As for transaction costs, workers knew that they would receive their full payment for the incentivized task in a single transaction. This should address concerns that bonuses and penalties are perceived as being enacted infrequently.

To check whether workers' beliefs about the task difficulty were affected by the

[10] Formally, if the participation constraint binds and the limited liability condition is slack, I can substitute for w in Π using the participation constraint. Differentiating the resulting expression with respect to F and using the fact that $\gamma < \bar{\gamma}$ and $b \leq v$ at the optimum reveals that Π is decreasing in F.

[11] Flat incentives are used to avoid workers being exposed to more than one form of incentive pay during the experiment.

[12] I cannot use the methodology of Karlan and Zinman (2009) here because it would expose workers to both frames and therefore likely make transparent the equivalence of the two.

[13] Bénabou and Tirole (2003) analyze an asymmetric information context whereby if the principal offers a larger bonus for a task (in equilibrium), the agent will believe the task to be more difficult. This kind of argument is difficult to formalize in the context of choice of contract frame; in Bénabou and Tirole (2003), the choice of incentive plays the role of a costly signal from principal to agent, whereas a frame in my context is pure cheap talk. While it may be possible to construct equilibria in which bonuses are taken to signal harder tasks, it is equally possible to construct the reverse equilibrium.

frame, I asked them at the beginning of stage 2 to estimate the average accuracy rate from stage 1. If workers who receive one contract perceive the task to be more difficult than those who receive a different one, they should estimate a lower mean performance from the first task.[14]

2.1 Experimental Platform: Amazon Mechanical Turk (MTurk)

The experiment was run on the online platform Amazon Mechanical Turk (MTurk, for short). MTurk is an online labor market for "micro outsourcing". For example, a "requester" that needs data entered, audio recordings transcribed, images categorized, proofreading, or many other possible types of tasks can post a Human Intelligence Task (HIT) on MTurk, and recruit "workers" to carry it out. Pay is set by the requester.

MTurk has many attractive features for research. For example, a short survey can be prepared, posted and completed by hundreds of workers in a matter of hours, typically for much smaller incentives than might be used in a laboratory experiment. Bordalo et al. (2012) test their theory of salience using MTurk surveys. Barankay (2011) uses MTurk to study the effect on willingness to undertake more work of telling workers about their rank in an initial task. Horton et al. (2011) and Amir et al. (2012) replicate some classic experimental results with MTurk subjects.

2.2 Effort task

In each stage of the experiment, subjects were asked to transcribe 50 text strings, gradually increasing in length from 10 characters to 55 characters. The strings were generated using random combinations of upper and lower case letters, numbers and punctuation and distorted to give the appearance of having been scanned or photocopied.

The task was chosen to be implementable online, to be reasonably similar to the types of tasks that participants are used to doing in the course of their work on MTurk, and to be sufficiently difficult to generate variation in performance (accuracy) without putting the workers under time pressure.[15] Time pressure was not used to maintain similarity with other MTurk tasks which typically allow workers to work in their own time. In each stage there were 10 possible sets of strings and participants were randomly assigned to one set.[16] An example screen is reproduced in Figure 2.

[14]Trust might also be important. To address such concerns, the design ensures that all participants have already interacted with me, the principal, through the first stage of the experiment. They have agreed to an informed consent form that states their work is part of a research project from an internationally well-known university (note that they were not told that it was an incentives study), gives my name and contact details. They were paid promptly after completing the first stage.

[15]The task closely resembles the kind of garbled text that individuals must type to solve a CAPTCHA (Completely Automated Public Turing test to tell Computers and Humans Apart) puzzle on the web. Such puzzles are used in web forms as an attempt to prevent bots and spammers from accessing sites; in fact this has led to some spammers recruiting MTurk workers to solve the CAPTCHAs that are blocking their access. See e.g. New York Times blog, March 13, 2008: http://bits.blogs.nytimes.com/2008/03/13/breaking-google-captchas-for-3-a-day/.

[16]This was done because other experimenters report occasionally participants posting answers to tasks on the web. I found no evidence of this occurring in this experiment.

2.3 Design specifics

A flowchart summarizing the design and timings is given in Figure 1. Two experimental sessions were conducted, each of which consisted of two stages. The first stage of the experiment recruited US-based workers on MTurk for a "Typing task and survey" for a flat pay of $3. Participants performed the typing task then filled out the survey, which is described below. Once all participants had been paid, they were sent an email informing them of their performance in the typing task.[17]

Six days later, all participants from stage 1 were sent a second email, inviting them to perform a new typing task, this time under experimentally varied incentives. Each contract has three components: a fixed pay component that does not depend on performance, a variable pay component that does depend on performance, and a frame that is either "bonus" or "penalty". Participants were told that the task would remain open for four days, and that they could only attempt the task once.

Performance pay was calculated as follows. Participants were told that after completion of the task I would select, using a random number generator, one of the 50 strings that they had been assigned to type, and that they would receive the bonus (avoid the penalty) conditional on that item being entered correctly. I avoided using emotive terms like "bonus" and "penalty". For example, a penalty framed offer in experimental session 1 was worded as follows: "The basic pay for the task is $3.50. We will then randomly select one of the 50 items for checking. If you entered it incorrectly, the pay will be reduced by $1.50." This particular pay structure was chosen because it means that the probability of receiving the bonus (avoiding the penalty) is equal to the accuracy rate (fraction of strings typed correctly).[18]

Participants were randomized into one of three possible financial incentives, and either bonus or penalty frame. The treatments are detailed in Table 1, and consist of either low fixed and variable pay, low fixed and high variable pay or high fixed and low variable pay. The choice of rates of pay is discussed in Appendix C.1.

Finally, in experimental session 2, participants were invited to a paid follow-up survey four days after stage 2 closed.

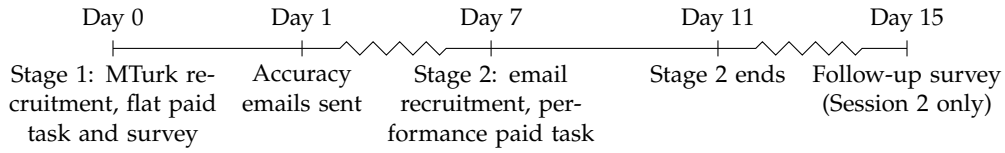


Figure 1: Experiment design flowchart.

[17] Example text is given in Appendix C.2.

[18] Examples of the full email text are given in appendix C.3. Experimental sessions 1 and 2 differed in the exact phrasing of the email, in order to check whether the results from Session 1 were driven by inattention.

Table 1: Treatments

		Contract[a]			Job offer[b]		
Group	N	Fixed pay	Variable pay	Frame	Base pay	Bonus	Penalty
Session 1							
0	192	\$0.50	\$1.50	Bonus	\$0.50	\$1.50	
1	188	\$0.50	\$1.50	Penalty	\$2		\$1.50
2	193	\$0.50	\$3	Bonus	\$0.50	\$3	
3	191	\$0.50	\$3	Penalty	\$3.50		\$3
4	193	\$2	\$1.50	Bonus	\$2	\$1.50	
5	189	\$2	\$1.50	Penalty	\$3.50		\$1.50
Session 2							
6	153	\$0.5	\$3.00	Bonus	\$0.50	\$3.00	
7	151	\$0.5	\$3.00	Penalty	\$3.50		\$3.00

[a] "Contract" details the three components of the contract offered: Fixed Pay (unconditional), Variable Pay (received if accuracy check is passed) and Frame.
[b] "Job offer" is the terms given in the email invitation to stage 2.

2.4 Data

This section describes the key variables collected in the survey and effort tasks. Summary statistics are given in Tables 2 and 3 and summary distributions of key variables plotted in Appendix Figure B.1.[19]

The measure of loss aversion I use is similar to that of Abeler et al. (2011), but unincentivized. Participants are asked to consider a sequence of 12 lotteries of the form "50% chance of winning \$10, 50% chance of losing \$X," where X varies from \$0 to \$11. For each lottery, they are asked whether or not they would be willing to play this lottery if offered to them by someone "trustworthy". I proxy for loss aversion with the number of rejected lotteries.[20] 7 percent of participants made inconsistent choices, accepting a lottery that is dominated by one they rejected. A screenshot of the lottery questions is given in Appendix C.5.[21]

Two other key variables that I attempt to measure are participants' reservation wages and their perceptions of what constitutes a "fair" wage. A measure of reservation wages is useful in considering how the framed incentives affect willingness to accept a job offer. All else equal (in particular, controlling for ability), if one contract is perceived as less attractive it should particularly discourage those with a higher reser-

[19] Participants were also asked to report the zipcode of their current location, which I map in Appendix B.11. The distribution of participant locations closely resembles the population distribution across the US.

[20] Note that by Rabin (2000) aversion to risk in small stakes lotteries is better explained by loss aversion than standard concave utility.

[21] The lottery choices were not incentivized because of concerns that this would interfere with studying selection effects and willingness to accept job offers for the effort task alone. Offering financial incentives large enough for participants to potentially lose \$10 is problematic because it would interfere with the selection effects I am trying to measure: if the incentives were advertised upfront they might attract high reservation wage participants who would not participate in the stage 2 effort task; if they were not pre-announced, subsequently revealing them might lead the participants to expect unannounced rewards in stage 2 and thus be more likely to accept in stage 2. Camerer and Hogarth (1999) argue that: "In the kinds of tasks economists are most interested in, like trading in markets, bargaining in games and choosing among gambles, the overwhelming finding is that increased incentives do not change average behavior substantively."

vation wage. To elicit reservation wages I ask participants what is the minimum hourly wage at which they are willing to work on MTurk.

Fehr and Gächter (2002) find in a buyer-seller experiment that penalty-framed performance incentives led to more shirking among sellers than equivalent bonus-framed offers, and argue that this is because the penalty contracts are perceived as less fair. I ask participants what they think is the minimum fair wage that requesters "should" pay on MTurk, and use this measure to proxy for fairness concerns. Reservation wages are typically lower than fair wages.

The main performance measure is "Accuracy Task X", the fraction of text strings that participants entered correctly in stage X. I also construct a second accuracy measure, "Scaled Distance Task X", which can be thought of as the error rate per character typed.[22] In the regressions I use the natural log of this measure since it is is heavily skewed by a small number of participants who performed poorly (per-string accuracy rates are sensitive to small differences in per-character error rates). I also try to measure how much time participants spent on their responses. There are large outliers since I cannot observe how long participants were actually working on a given page of responses, only how long the page was open for, so I take the time the participant spent on the median page, multiplied by 10 to estimate the total time. Finally, at the beginning of stage 2 participants were asked to estimate the mean accuracy rate from stage 1, a variable I label "Predicted Accuracy".

In total 1,465 participants were recruited, of which 693 returned for stage 2. 15 participants are dropped from all of the analysis, six because I have strong reasons to suspect that the same person is using two MTurk accounts and participated twice and nine because they scored zero percent accuracy in the stage 1 typing task, suggesting that they did not take the task seriously (of the six of these who returned for stage 2, five scored zero percent again).

2.5 Randomization

I stratified the randomization on the key variables on which I anticipated selection: stage 1 performance, rejected lotteries and reservation wage. I was a little concerned that some participants might know one another (for example, a couple who both work on MTurk), so the treatments were randomized and standard errors clustered at the zipcode-session level.[23] In regressions that drop participants who share a zipcode this is equivalent to using robust standard errors, since then each cluster is of size one.

As a graphical check of balance, Appendix B.3 plots the CDFs for task 1 accuracy, reservation wage, fair wage and rejected lotteries, separately for the bonus and penalty

[22] For each text string I compute the Levenshtein distance between the participant's response and the correct answer, and divide by the length of the correct answer. The Levenshtein distance between two strings, A and B, is the minimum number of single character insertions, deletions, or swaps needed to convert string A into string B. This then roughly corresponds to the probability of error per character for that string. I then take the average over all text strings for that participant to find their per-character average error rate

[23] In the stage 1 data 187 individuals report being located in the same zipcode as another participant from the same session.

treatments, confirming good balance on these variables.[24]

Table 4 gives the results of the statistical balance tests. I perform two exercises. The first tests the joint significance of the full set of treatment dummies in explaining each baseline characteristic. The second performs a t-test for comparison of means between pairs of treatments, where each pair considered differs only in terms of its bonus/penalty frame (groups (0,1), (2,3), (4,5) and (6,7) as labeled in Table 1). Both exercises confirm good mean balance on all characteristics with the exception of the minimum fair wage (F-statistic p-value 0.01), where the difference comes from differences between sessions 1 and 2, and the number of MTurk HITs completed (p-value 0.07), where the difference is driven by a small number of participants with very large numbers of HITs completed.[25]

3 Main Results

This section discusses the effect of the penalty frame on participants' willingness to accept the contract, on the types of participants who select into the contract, and on performance on the job. I discuss the relation between the key observable characteristics, and between characteristics stage 1 performance, in Appendix B.2.

3.1 Acceptance

Figure 3 graphs the rates of acceptance of the stage 2 job offer by treatment. The striking pattern in these data is that penalty framed contracts were much more likely to be accepted than equivalent bonus framed contracts. The relationship is pronounced for the four groups with fixed pay of \$0.50, and weaker for the two groups with fixed pay equal to \$2. In addition, acceptance is substantially higher under higher fixed pay, while the relation between variable pay and acceptance appears weak at best.

This result is particularly notable because it directly contradicts model Prediction 3. The model predicts that penalty contracts should be unattractive relative to equivalent bonus contracts. I discuss this finding in relation to the theory in section 5.

The basic regression specification is a linear probability model with dependent variable $Accept_i \in \{0, 1\}$, individuals indexed by i:

$$Accept_i = \beta_0 + \beta_1 * Penalty_i + \beta_2 * HighFixed_i + \beta_3 * HighVariable_i + \beta_4 * X_i + \epsilon_i \quad (5)$$

Penalty is a dummy equal to 1 if the contract is penalty framed and zero if bonus framed. *HighFixed* is a dummy indicating fixed pay equal to \$2 (alternative: \$0.50). *HighVariable* is a dummy indicating variable pay of \$3 (alternative: \$1.50).[26] X_i is a

[24] Appendix B.4 plots the distributions of these variables by experimental session.

[25] In addition, I run Kolmogorov-Smirnov or Mann-Whitney equality of distributions tests between bonus and penalty frames for stage 1 performance, time spent on stage 1, loss aversion, reservation wage and fair wage, none of which reject the null of equal distributions.

[26] Since there are only two levels of fixed and variable pay, it is straightforward to compute the implied linear effects (per dollar), by dividing the coefficient on *HighFixed* by 1.5 (\$2 - \$0.50) and the coefficient on *HighVariable* by 1.5 (\$3 - \$1.50).

vector of variables measured in stage 1. In particular, I include accuracy and time spent on the stage 1 effort task, to jointly proxy for ability and intrinsic motivation. All results are robust to additionally including the ratio of accuracy to time spent (not reported). X_i also includes dummies for the set of items assigned to be typed by that participant (10 possible sets per stage). Note that the main specifications pool the effects of each component of the contract to increase power.

Table 5 presents the main results. I find that a penalty framed contract increases acceptance rates by approximately 10 percentage points over the equivalent bonus frame. This implies a 25 percent higher acceptance rate under the penalty frame than the bonus frame (the mean acceptance rate under the bonus frame was 42 percent), which seems a large effect for a simple framing manipulation. High fixed pay increases acceptance by around 15-16 percentage points. Surprisingly, the effect of high variable pay is positive but much smaller at around 3 percentage points greater take-up, and not statistically significant. The results are robust to dropping participants who made inconsistent choices in stage 1, who spent a very long time on the first task, have very high reservation or fair wages, or are from zipcodes with more than one respondent. Near-identical average marginal effects are obtained using logistic instead of linear regression.

Participants who performed better on the unincentivized stage 1 were significantly more likely to accept the stage 2 job offer, as is clear from Table 2 and Figure B.5. This is consistent with the common finding that performance pay differentially selects more able or motivated workers, and which I discuss further in Appendix B.10. Participants with a higher reservation wage were significantly less likely to accept the offer. When controlling for this measure, the coefficient on "minimum fair wage" is not statistically significant, suggesting that fairness concerns (as measured by this variable) were not of primary importance for willingness to accept the contract.

The number of hypothetical lotteries rejected by participants is not predictive of acceptance, whether or not I drop participants who made inconsistent choices in the lottery questions. This is surprising as the stage 2 contract is risky, so one would expect more risk/loss averse participants to be less willing to accept it. Appendix Figure B.5 shows that the distributions of rejected lotteries are essentially identical for participants who did and did not accept the stage 2 job offer. This could be because the measure is poorly capturing loss aversion, although similar unincentivized measures have been successful in other studies (see Camerer and Hogarth (1999)).

Visual inspection of Figure 3 suggests that the effect of the penalty frame is larger when the variable pay component is larger and smaller when the size of the fixed pay component is larger. Table 6 reports the relevant interaction effect estimates. The point estimates do indeed suggest that the effect of the penalty frame is smaller for high fixed pay and larger for high variable pay, however neither estimate is statistically significant when estimated separately or simultaneously. Note that this does not imply that the framing effect will disappear for, for example, bigger jobs with higher total pay. It is entirely consistent with the fact that acceptance rates must converge as we move into the right tail of the reservation wage distribution, by increasing the fixed pay component.

In addition, the point estimate on "high variable pay" is essentially zero for participants under the bonus frame, implying that the potential for a $3 bonus as opposed to a $1.50 bonus did not make the job offer significantly more attractive.

3.2 Selection

Now I turn to the effect of the penalty frame on the types of workers that select into the contract. Figure 4 plots CDFs of stage 1 task performance, time spent on stage 1 task, rejected lotteries, reservation wage and fair wage, comparing those who accepted the bonus frame with those who accepted the penalty frame. Surprisingly, the distributions are overlapping for all variables except for reservation wages, implying no notably differential selection on these variables.[27] I do observe suggestive evidence that the penalty contract attracted workers with higher reservation wages. This is not inconsistent with no selection on other characteristics, since as shown in Appendix B.2, the correlation between reservation wages and other characteristics is small.

Table 7 tests for selection effects of penalty framing by interacting the penalty frame with the key observables in acceptance regressions. The interaction coefficients estimate the extent to which a given characteristic more or less strongly predicts acceptance under the penalty frame. In each case the interaction terms are not statistically significant, whether estimated separately or jointly. A joint test fails to reject the null that all interaction coefficients are equal to zero (p-value 0.90).

In Appendix B.5 I check if the results are robust to dropping outliers for time spent on task 1, reservation wage and fair wage, or dropping participants who made inconsistent lottery choices. The only difference is that now the interaction between penalty frame and reservation wage is significant at the 10 percent level, consistent with the penalty screening in participants with higher reservation wages. Overall there is little evidence of selection on these key observables between contract types.

As for the other covariates, men and women are equally likely to accept the penalty contract, but men are eleven percentage points less likely to accept the bonus contract. In other words, men are less likely than women to accept the incentivized task in general, but relatively more likely to prefer the penalty to the bonus contract. Participants who reported that their main reason for working on MTurk is to earn money (93 percent of participants) are relatively more likely to accept the penalty contract than those who gave another reason, similar applies to those who report mostly working on research HITs. Participants with more HITs completed are relatively less likely to accept the bonus contract. Despite the fact that being male and citing money as the main reason for working are positively associated with performance on stage 1, none of these results seems to be consequential for performance, as illustrated by the lack of selection on stage 1 performance measures and the evidence presented in the next section.

It is quite surprising that there seems to be no selection effect of penalty framing.

[27] The equivalent "balance" figures in Appendix B.3 confirm that this finding represents balance before and after candidates accepted the offer, not an initially unbalanced situation that became balanced by chance after acceptance decisions.

One possibility is that selection is hard to detect in this context. If workers could choose between the bonus and penalty framed job (assuming this did not undo the framing effect) they should select into the job they preferred. However in this experiment, workers cannot choose jobs, only whether to accept the one that they are offered. As a result there will be many participants in each pool whose participation constraints would be satisfied under either contract: any selection effect would have to be driven by the fraction of participants whose participation constraints are satisfied under one but violated under the other. However, given the large difference in acceptance rates between contracts it seems unlikely that this is what is driving the lack of detectable selection. In addition, as documented in Appendix B.10, I am able to observe the more standard result of differential selection by type into the incentivized task: workers with high performance in the first stage are more likely to accept the job in the second stage.

3.3 Performance

Now I turn to the incentive effects of contract framing on worker effort and performance. This section directly relates to the existing literature on framed contracts which considers incentive effects for an already recruited sample of workers or participants.

The basic regression equation is:

$$Y_i = \delta_0 + \delta_1 * Penalty_i + \delta_2 * HighFixed_i + \delta_3 * HighVariable_i + \delta_4 * X_i + \epsilon_i \quad (6)$$

Where Y_i is a measure of effort or performance. The key measures are summarized in Table 2 and distributions plotted in Appendix Figure B.1. As before, X_i is a vector of variables measured in stage 1.

In general one would expect the estimates of δ_1, δ_2 and δ_3 to be biased by selection: if the workers that accept one type of contract are different from those that accept another, then performance differences may simply reflect different types rather than different effort responses to incentives. However as already documented, I do not observe notably differential selection between frames, which would bias the estimate of the key coefficient of interest, δ_1. Moreover, since I have stage 1 measures of performance and characteristics, I can control for selection on observables by including these.

Figure 5 presents the mean performance on the stage 2 task by treatment group. I find that performance is higher under the penalty than under the bonus frame, consistent with the existing experimental studies. Pooling each framing treatment, in Figure 6 I plot CDFs of the accuracy measure, the log distance measure (recall that this is interpreted as the log of the per-character error rate) and time spent, and find that performance and effort is higher under the penalty frame right across the distribution.

The main regression results are given in Table 8. I find that accuracy under the penalty frame is around 3.6 percentage points (around 0.18 standard deviations or 6 percent of the mean accuracy of 0.59) higher than under the bonus frame, statistically significant at 5 percent without and 1 percent with controls. The coefficient estimate is robust to dropping participants who made inconsistent lottery choices, participants

from zipcodes with multiple respondents, and outliers on the reservation and fair wage questions (although a little smaller and only significant at 10 percent). Crucially, the estimated penalty effect is unaffected by the inclusion or exclusion of controls, consistent with the contract frame not inducing significant outcome-relevant selection, at least on observables. For selection to explain the results, there would have to be a substantial unobserved driver of performance that differentially selected under the penalty frame and orthogonal to the set of controls included in the regressions.[28]

In addition, high fixed pay increases accuracy by around 2-4 percentage points, significant at 5 percent when controls are included. The point estimate doubles when controls are included, suggesting there may be adverse selection induced by the higher fixed pay. If anything, the fact that a selection effect *is* observed here gives comfort that the lack of observed selection between bonus and penalty reflects a true lack of selection in the data. Lastly, high variable pay increases accuracy by around 1.4-2.5 percentage points, although this is never significant at conventional levels.

As for the other key variables, performance in the first stage very strongly predicts performance in the second stage, while the coefficient on time spent in the previous task is negative, small in magnitude and not significant. As in stage 1 (see Appendix B.2), a higher reservation wage is associated with poorer performance. Controlling for the reservation wage, the reported minimum fair wage has no effect on performance, consistent with fairness concerns not being of primary import.

In this stage the number of rejected lotteries is negatively associated with performance, significant at ten percent when including participants who made inconsistent choices and five percent when they are dropped. A one standard deviation increase in the number of rejected lotteries is associated with around 1-2 percentage points worse performance. This result relates to Prediction 2 and is expanded upon when I consider heterogeneous effects below.

In Table 9 I separate the effects of increasing fixed and variable pay under bonus and penalty framing. Increasing the fixed pay seems to have the same effect under both bonus and penalty contracts (the interaction effect is an imprecisely estimated zero). Increasing the size of the variable pay is associated with higher effort under both frames, with a smaller effect under the penalty frame. However neither estimate is statistically significant.

Table 10 reports estimates of heterogeneous effects of the penalty treatment by the main variables. In each case the individually estimated interaction effect is not statistically significant: there is little evidence of strong heterogeneous effects.[29]

[28] In Appendix table B.4 I regress the distance measure of accuracy (log errors per character typed) and time spent on treatment dummies and covariates. The estimates imply that the penalty frame led to participants committing 20 percent fewer errors per character (from a mean of 0.066), and spending around two to three minutes longer on the task (mean 41 minutes), although the latter is not significant when controls are included. The point estimates on fixed and variable pay mirror their counterparts in the main regressions, and once again there is evidence of adverse selection induced by the higher fixed pay.

[29] Including all of the interaction effects estimated together gives one unexpected and difficult to interpret result: a negative interaction effect for reservation wages and a positive one for minimum fair wages. Dropping participants above the 99th percentile for reservation or fair wages, the p-value on the reservation wage interaction increases to 0.09 and that on the fair wage interaction increases to 0.14.

Focusing on the coefficient on rejected lotteries, I note that although neither the main effect nor interaction coefficient are statistically significant, nevertheless it is striking that the implied coefficient on rejected lotteries is close to zero under the bonus frame, and negative under the penalty frame (the combined effect under the penalty frame is statistically significant at the 5 percent level), while the model Prediction 2 implies that the coefficient should be more positive under the penalty frame. Note also that the model predicts a positive relationship between loss aversion and performance, whereas I find a negative one. An extension that allows the reference point to also depend upon expectations, outlined in Appendix A.3, can allow for this finding.

I lack power to dig into this relationship in depth. I do however perform one simple exercise. In Figure 8 I non-parametrically plot accuracy against rejected lotteries separately under bonus and penalty frame, after partialling out the other variables, dropping participants with inconsistent choices and those who rejected or accepted all lotteries. The slopes are approximately equal over much of the range of values for rejected lotteries, but flattening out for high values under the bonus frame while becoming strongly negative under the penalty frame, which seems to be what is what is driving the difference in the regression coefficients. For most participants the relationship between performance and loss aversion is similar between frames, but penalties seem to strongly *discourage* the most loss averse participants, an effect which is not in the model.

The surprising implication of the results is that a simple switch from bonus to penalty framing is very lucrative from the principal's perspective. My estimates suggest that recruitment and performance would be approximately equal under a contract that pays \$0.50 fixed pay with \$1.50 variable pay framed as a penalty, as with \$1.50 fixed pay and \$1.50 variable pay, framed as a bonus.

4 Secondary Results

In this section I discuss secondary results aimed at partly unpacking the mechanism behind the main results. First, I demonstrate that the penalty frame did not appear to change participants' perceptions of the task difficulty (I find the same result in the follow-up survey described in section 5.1). Second, I argue that inattention is unlikely to explain the higher acceptance rate of the penalty contract.

Additionally, in Appendix B.9, I discuss persistence, showing that the performance difference between bonus and penalty frame persisted throughout stage 2. If workers quickly realized the equivalence of the frames, one would expect performance to converge. In Appendix B.10 I document that a standard selection result is present in my data: performance pay attracted more able workers, which partly explains the difference in stage 1 and stage 2 performance. This gives confidence that the no-selection result between frames is not driven by, for example, MTurk workers being atypical experimental subjects.

4.1 Does framing affect perceived task difficulty?

At the beginning of the stage 2 task, subjects were asked to predict the overall mean accuracy rating from the first stage. If the frame influenced their beliefs about the task difficulty, it should also influence their belief about mean performance in the first stage. Figure 7 graphs the mean of these predictions by treatment group. There is no systematic relationship between the framing treatment and the predictions.

In Appendix B.7 I regress the participants' estimates on the treatment and the same controls as the main regressions. As expected, I find no evidence that the frame influenced participants' beliefs about the task difficulty: I can rule out a difference larger than ± 3 percent at the 95 percent level. In sum, it does not appear that the effect of the penalty contract is explained by different beliefs induced by the frame.[30]

4.2 Inattention: Comparing Sessions 1 and 2

After running the first experimental session, one concern was that the way in which the job offer was phrased might differentially attract inattentive participants into the penalty frame, explaining the higher acceptance rate. The "pay" section of the offer email under the bonus (penalty) frame was written as follows:

> The basic pay for the task is \$0.50 (\$3.50). We will then randomly select one of the 50 items for checking. If you entered it correctly (incorrectly), the pay will be increased (reduced) by \$3.00.

It is possible that an inattentive participant receiving the bonus email might glance at the first sentence, see a low amount and close the message, while under the penalty frame she might see the high amount and click through to the task.[31] To alleviate this concern, the second session changed the email slightly, as follows:

> The pay for this task depends on your typing accuracy. We will randomly select one item for checking, and if it was entered correctly (incorrectly), the pay will be increased above (reduced below) the base pay. The base pay is \$0.50 (\$3.50) which will be increased (reduced) by \$3 if the checked item is correct (incorrect).

This phrasing pushes the pay information to the end of the paragraph and puts it all into one sentence. It also tells the participants immediately that the pay will depend on performance, and hence they should pay attention to the contingent component of pay.

[30] I do find that participants assigned the high bonus predict that stage 1 accuracy was 3.5 percentage points higher. To check whether this affected stage 2 performance, I also regress stage 2 performance on the main regressors, now including the participant's prediction of stage 1 accuracy. In the specification with only treatment indicators I find a strong and significant correlation between predicted and actual performance, perhaps because the predictions are positively (although not significantly) correlated with own performance on task 1, which is in turn a strong predictor of task 2 performance. Taken literally, the coefficient implies that a 10 percentage point higher belief about mean stage 1 performance is associated with 2 percentage points better performance in stage 2. When controls are included, this coefficient drops effectively to zero and is not statistically significant.

[31] Note, however, that this is only part of the email content and that "base pay" should strongly signify that there is more payoff-relevant information to come.

Of course, one challenge with interpreting such manipulations is that if no effect were observed, it is hard to tell whether the rephrasing eliminated inattention, or somehow changed the reference point.

As shown in Figure 3, the penalty framed offer was again strongly more attractive than the bonus framed offer in session 2, with an acceptance rate of 48 percent under the penalty and 37 percent under the bonus, suggesting that inattention does not explain the main results. I do find that the performance difference between bonus and penalty was much smaller and not significant in session 2 (see Figure 5). I confirm in this by splitting the penalty treatment effect by session in regressions in Appendix B.8. It is not possible to say whether this is a treatment effect (the rephrasing decreased the framing effect on performance) or sampling variation.

5 Why are penalties popular?

The acceptance rate is substantially higher under the penalty frame than under the bonus frame, contradicting model Prediction 3. While loss aversion can explain the higher effort provision under the penalty contract, it cannot explain the higher acceptance rate, suggesting that participants may be maximizing a different objective function when deciding whether to accept the job than when performing the task.[32]

Participants appear to be failing to anticipate the loss they will experience if unsuccessful, should they take accept the job.[33] It could be that the temporal separation between doing the task and realization of the payment a few days later is sufficient that participants do not anticipate being strongly disappointed when unsuccessful under the penalty frame. However, clearly this is not enough to explain the higher acceptance rate under the penalty contract, since complete failure to anticipate loss aversion should imply indifference between the two contracts.

An obvious candidate explanation is that participants anticipate they will work harder and earn more under the penalty contract, which makes it attractive. This argument requires three parts. First, that the loss when unsuccessful is not anticipated as being sufficiently painful to discourage acceptance, for instance because the "planner" self who accepts the offer is not averse to losses experienced by the "doer" self who does the work. Second participants must recognize that they will work harder under the penalty frame. Third, effort provision under the bonus frame must be suboptimally low, such that being motivated to work harder is attractive. In other words, it requires a self-control problem, where penalties act as a commitment device.[34]

As a simple example, suppose that when deciding whether to accept the contract,

[32] As such, the following discussion relates to multiple-selves models e.g. Thaler and Shefrin (1981), Fudenberg and Levine (2006).

[33] This point relates to the literature on anticipating preferences or biases. For example, Loewenstein and Adler (1995) find that people underestimate the endowment effect: they predict a lower willingness to accept to give up a mug when asked to imagine being endowed with it than when actually endowed with it. See also Loewenstein et al. (2003) for related discussions.

[34] This point is closely related to the goal-setting literature, particularly Koch et al. (2012) and Golman and Loewenstein (2012).

A evaluates it according to the following modified utility function:

$$V(w,b,e^*(w,b,F)) = w + 2e^*(w,b,F)b - \beta\frac{e^{*2}(w,b,F)}{2\gamma} + \alpha e^*(w,b,F). \tag{7}$$

(7) is chosen to coincide with (1) when $\lambda = \beta = 1$ and $F = 0$, in other words, when A does not expect (or is not averse to) losses. V depends on $e^*(w,b,F)$, reflecting that A knows that if she accepts the contract she will choose her effort according to (3). A prefers the penalty frame ($F = 1$) if $V(w,b,e^*(w,b,1)) > V(w,b,e^*(w,b,0))$.

From the perspective of an agent maximizing V, the first-best effort choice is $e^{FB} = \frac{\gamma[\alpha+2b]}{\beta} = \frac{e^*(w,b,0)}{\beta}$. It is clear therefore that a necessary condition for A to prefer penalty frames is $e^{FB} > e^*(w,b,0)$, or $\beta < 1$. One interpretation of this condition is that A has weak self-control and is sophisticated about it: she knows she will exert less effort than she would like to. A simple sufficient condition for this self-control problem to be sufficiently severe that the penalty is preferred is $e^{FB} \geq e^*(w,b,1)$ or $\frac{\alpha+2b}{\alpha+b(1+\lambda)} \geq \beta$.

It is certainly plausible that workers might like penalty contracts as a commitment device, an idea reminiscent of recent evidence from Kaur et al. (2013), who find that workers select into a strictly dominated financial incentive scheme that acts as a form of commitment to higher effort provision. However, in my context it seems unlikely that this is what is driving the large differences in acceptance rates between bonus and penalty frames, because the associated earnings difference is small (not even accounting for the higher effort cost under penalties). In Table 11 I compute what would be the average rational expectation of earnings for each treatment, computed as fixed pay plus the product of mean task 2 accuracy and variable pay. The increase in expected earnings when switching from bonus to penalty framing is never more than 10 cents, but generates around 10 percentage points higher acceptance. Meanwhile increasing the fixed pay by \$1.50 (which increases expected pay by just over \$1.50 after accounting for effort) increases acceptance by 19 percentage points. As previously noted, increasing the variable pay under the bonus frame has no appreciable effect on acceptance, despite increasing expected earnings by around 90 cents.

These results suggest that workers do not have rational expectations about their earnings, or are somehow failing to correctly weight the terms of the contract. The most likely explanation is that participants are focusing on the "base pay", which is higher under the penalty contract, and underweight the bonus or penalty component. Thus the base pay both shifts their reference point (driving differences in effort) and their valuation of the contract. A simple reduced form way to model this is to simply assume that the frame, F, enters additively on the left hand side of A's participation constraint.

Participants may be evaluating the terms of the contract separately, instead of integrating them into an expectation. In other words, they are acting as if they have preferences over base pay and contingent pay, rather than over outcomes. The contract (w,b,F) is perceived as $(\text{base pay}, \text{bonus}, \text{penalty}) = (w + Fb, (1-F)b, Fb)$ such that $(w+b,0,-b) \succ (w,b,0)$, rather than as a lottery that pays $w+b$ with probability

e and w with probability $1-e$. Essentially, base pay is overweighted when evaluating whether to accept: $(\$3.50, \$0, -\$3)$ feels more lucrative than $(\$0.50, \$3, \$0)$. This possibility is closely related to the finding of Hossain and Morgan (2006), that in online auctions for identical goods, the total sale price (winning bid plus shipping cost) is higher when the shipping cost is higher (revenue equivalent predicts that the total price should not depend the proportion labeled as shipping costs). Hossain and Morgan (2006) suggest a mental accounting explanation, whereby goods and shipping are assigned separate mental accounts and are subadditively weighted.

The findings might also reflect a form of "Coherent Arbitrariness" (Ariely et al. (2003)). An agent with coherently arbitrary preferences is vulnerable to arbitrary manipulations of her valuation of a good or experience (for example, her valuation of a good can be altered by priming her with a transparently uninformative random number). In this context, participants' valuations of the contract are influenced by the number presented as base pay, which is higher under the penalty contract.

A third mechanism that would manifest as an apparent preference for high base pay is analyzed theoretically by Just and Wu (2005) and Hilken et al. (2013). In their models, the contract and outside option are both evaluated against the same reference point, which is influenced by the frame. By increasing the reference point, the penalty frame makes the contract appear more attractive relative to the outside option.

5.1 Evidence

To shed some additional light on mechanisms, I conducted a short survey four days after experimental session 2 ended, asking participants their perceptions of the job offers. The questions are unincentivized and mostly subjective, and the survey was conducted after completion and payment of stage 2 which might affect responses, so the results should be taken with a large pinch of salt. All participants from stage 1 of session 2 were invited to complete a survey for a fixed incentive of \$2. 82 percent did, balanced between the bonus and penalty frame.[35]

Participants were first reminded of the job offer they received, then asked a series of questions about it. Appendix table B.8 presents results. Participants were asked to indicate agreement on a 1-7 scale to whether their job offer or task was fun, easy, paid well, fair, was a good motivator, whether the principal could be trusted, achievable[36] and understandable (Panel A). They were then asked to what extent they agreed that various features made the offer attractive or unattractive (Panel B). Third, they were asked to guess mean acceptance and performance of participants who received the same job offer as they did.

For most questions I find no significant differences between frames. However the penalty offer was rated significantly higher (about 0.3 s.d., significant at 5 percent)

[35] Participants who completed stage 2 were more likely to complete the survey (96 percent vs 73 percent, p-value < 0.001), probably reflecting that some non-participation in stage 2 is driven by participants who did not see my emails.

[36] "If a participant worked hard on the task, he or she can be confident that they would answer the checked item correctly."

for "good pay" and was more likely to be considered attractive due to good pay. If anything, the penalty was perceived as a less good motivator and less attractive for its motivational power than the bonus contract.

The second finding is that estimated acceptance rates and performance were not significantly different between bonus and penalty frames. It is possible that although the penalty increases a participant's own valuation and effort, they fail to recognize it will have this effect on others. Of course an alternative explanation is that participants were simply unable to make an informed guess.[37]

Overall I interpret the survey results as suggestively supporting the idea that the penalty contract, via its high "base pay", was viewed as more lucrative. They do not appear to support penalties being valued as a commitment device, although note that this reasoning is more complex to articulate and so perhaps harder to detect via survey questions.

5.2 Discussion

Although a demand for commitment story or overweighting of the base pay may explain why penalties were more popular, four findings remain outside the model and would benefit from further theoretical and empirical work.

First, participants exert higher effort when the fixed pay component is increased. The most obvious explanation here is reciprocity in response to perceived generosity from the principal (increasing the fixed pay).[38]

Second, it is surprising that a $1 increase in the variable pay seems to have at most a small and not statistically significant effect on acceptance and performance, while a $1 increase in the base pay (with no change in material incentives) has a large effect.

Third, given the size of the effect of the penalty frame on acceptance rates, it is very striking that there is apparently no selection effect of incentive framing on the observables measured. Almost any model is likely to predict adverse or advantageous selection. It would be particularly interesting to see whether this finding replicates in other settings.

Fourth, the relationship between incentivized performance and loss aversion, measured by rejected lotteries, remains something of a puzzle. First, I find a negative relationship, which is inconsistent with the model outlined in section 1 but is possible

[37] At the end of the survey, I also presented participants with the alternative contract frame that they could have received and asked them to compare it to the one they did receive on generosity, motivational power, et cetera. Unsurprisingly, I find no significant differences between frames, consistent with subjects realizing the equivalence of the two. I also asked subjects how likely they would be, on a 5 point scale, to accept the job if it were offered to them again, to get a sense of how penalty contracts might perform in a repeated contracting environment. I find no significant difference in willingness to re-accept between bonus and penalty contract. Conditional on stage 2 performance, workers who were lucky (the item I checked was one of the ones they answered correctly) reported 0.6s.d. higher "likelihood".

[38] Alternatively, perhaps some participants believed that maintaining a good reputation with the principal would be rewarded in future, and higher fixed pay increased the perceived value of a good reputation. Participants were explicitly told in the first stage that the chance of being invited for future tasks would not depend on their performance, but this was not reiterated in the second stage. To explain the higher effort in all treatments under the penalty contract this reputation mechanism would have to be stronger for penalties than bonuses.

under a generalization that also allows the reference point to depend upon expectations, given in Appendix A.3. However, both predict a relatively more positive relationship under the penalty frame, which I do not observe. I presented evidence (albeit suggestive at best) that penalties may have a discouragement effect on the most loss-averse participants. This again would benefit from further research to unpack how loss aversion drives the effect of penalties.

6 Conclusion

This paper analyzes the effects of framed incentive pay on worker recruitment, selection and performance. I find that penalty framed incentives increased the number of workers who accepted the job by 25 percent relative to economically equivalent bonus framed incentives. In addition the penalty frame increased performance on the job by 6 percent, around 0.2 standard deviations. The effect sizes are large relative to those for standard manipulations of incentive size: increasing the non-contingent pay from \$0.50 to \$2 increased the acceptance rate by 36 percent and performance by 0.2 standard deviations, while the effect on both outcomes of increasing the contingent component of pay from \$1.50 to \$3 was small and statistically insignificant. Interestingly, I find no evidence of either adverse or advantageous selection into the penalty framed contract, while I do find advantageous selection into the incentivized task as a whole.

I present further evidence that the relative attractiveness of the penalty frame is not driven by changed perceptions of the difficulty of the task, nor is it driven by the wording of the job offer that might differentially attract inattentive workers. I also show in the appendix that a standard selection result obtains in this context, namely that the job offer with incentive pay attracted relatively high ability participants, and that the effect of the framed incentive on performance did not wear off during the course of the experiment.

While loss aversion, combined with an assumption that penalty framing increases the agent's reference point, predicts higher effort under the penalty frame, it also predicts that agents should be less willing to accept a penalty framed contract than the equivalent bonus contract, because agents anticipate that they will be more disappointed under the penalty contract. I propose two possible extensions that might explain why I find the opposite. One possibility is that agents like penalties because they motivate them to work harder, overcoming a self control problem, in a similar spirit to recent findings by Kaur et al. (2013). Alternatively, workers may be evaluating contract offers as if they had preferences over their constituent components (base pay, bonus, penalty), rather than over outcomes, and may be attracted by the salient high base pay under the penalty contract, similar to how eBay bidders apparently fail to correctly combine the cost of shipping and the price of the good when bidding (Hossain and Morgan (2006)). The fact that earnings were in practice very similar between bonus and penalty participants seems to go against them being particularly valuable for overcoming weak self-control. Meanwhile, a follow-up survey found that participants perceived

the penalty offer as more generous, suggestively supporting the second mechanism.

Overall, the results are surprising, suggesting that from the principal's perspective, penalty contracts may strongly outperform bonus contracts in some settings. In environments similar to the experimental context, such as short-term one-off engagements, firms may be able to gain through greater use of penalties. However to answer the motivating question of why firms seem reluctant to use penalties in general, it is clear that more research is needed.

The results and proposed mechanisms suggest that the answer may lie in dynamics. In practice, many contracting arrangements are repeated or long-term, and the contract must satisfy not only a participation constraint at the time of acceptance, but on an ongoing basis, to prevent workers from quitting. Meanwhile, it seems reasonable to expect that framing effects will eventually wear off: the agent's reference point adjusts, moderating the commitment power of the contract, or her rosy perception of the job offer adjusts over time to reflect what she sees on her payslips. Thus the higher acceptance rate under penalty contracts may also lead to a higher quit rate, which is costly for the employer.[39] Moreover, as the worker's reference point adjusts, the effect of the frame on her effort wear off as well, eliminating the contract's performance advantages.[40] Such effects may partly explain why firms are reluctant to use penalty frames.

References

Abeler, J., A. Falk, L. Goette, and D. Huffman (2011). Reference Points and Effort Provision. *American Economic Review 101*(2), 470–492.

Amir, O., D. G. Rand, and Y. K. Gal (2012, January). Economic games on the internet: the effect of $1 stakes. *PloS one 7*(2), e31461.

Ariely, D., G. Loewenstein, and D. Prelec (2003, February). "Coherent Arbitrariness": Stable Demand Curves Without Stable Preferences. *The Quarterly Journal of Economics 118*(1), 73–106.

Armantier, O. and A. Boly (2012). Framing of Incentives and Effort Provision. *mimeo*.

Baker, G., M. Jensen, and K. Murphy (1988). Compensation and Incentives: Practice vs. Theory. *The Journal of Finance 43*(3), 593–616.

Barankay, I. (2011). Rankings and Social Tournaments: Evidence from a Crowd-Sourcing Experiment. *mimeo*, 1–30.

Bénabou, R. and J. Tirole (2003, July). Intrinsic and Extrinsic Motivation. *The Review of Economic Studies 70*(3), 489–520.

Bordalo, P., N. Gennaioli, and A. Shleifer (2012, April). Salience Theory of Choice Under Risk. *The Quarterly Journal of Economics 127*(3), 1243–1285.

Brooks, R. R. W., A. Stremitzer, and S. Tontrup (2013). Stretch It but Don't Break It: The Hidden Risk of Contract Framing. *Working paper*.

[39] Salant and Siegel (2013) discuss this mechanism in the context of sales and returns of goods.

[40] I do not see such effects over the brief duration of my study (see Appendix B.9), but Jayaraman et al. (2014) do see only short-lived "behavioral" responses to a contract change.

Camerer, C., L. Babcock, G. Loewenstein, and R. Thaler (1997, May). Labor Supply of New York City Cabdrivers: One Day at a Time. *The Quarterly Journal of Economics 112*(2), 407–441.

Camerer, C. and R. Hogarth (1999). The effects of financial incentives in experiments: A review and capital-labor-production framework. *Journal of Risk and Uncertainty 19*(1-3), 7–42.

Church, B. K., T. Libby, and P. Zhang (2008, December). Contracting Frame and Individual Behavior: Experimental Evidence. *Journal of Management Accounting Research 20*(1), 153–168.

Crawford, V. P. and J. Meng (2011, August). New York City Cab Drivers' Labor Supply Revisited: Reference-Dependent Preferences with Rational-Expectations Targets for Hours and Income. *American Economic Review 101*(5), 1912–1932.

de Meza, D. and D. Webb (2007). Incentive design under loss aversion. *Journal of the European Economic Association 5*(March), 66–92.

Dohmen, T. and A. Falk (2011, April). Performance Pay and Multidimensional Sorting: Productivity, Preferences, and Gender. *American Economic Review 101*(2), 556–590.

Druckman, J. N. (2001, April). Using Credible Advice to Overcome Framing Effects. *Journal of Law, Economics, and Organization 17*(1), 62–82.

Eriksson, T. and M. C. Villeval (2008). Performance-pay, sorting and social motivation. *Journal of Economic Behavior & Organization 68*(2), 412 – 421.

Farber, H. S. (2005, February). Is Tomorrow Another Day? The Labor Supply of New York City Cabdrivers. *Journal of Political Economy 113*(1), 46–82.

Farber, H. S. (2008, May). Reference-Dependent Preferences and Labor Supply: The Case of New York City Taxi Drivers. *American Economic Review 98*(3), 1069–1082.

Fehr, E. and S. Gächter (2002). Do Incentive Contracts Undermine Voluntary Cooperation? *IEW Working Paper No. 34*.

Fryer, R. G., S. D. Levitt, J. List, and S. Sadoff (2012). Enhancing the Efficacy of Teacher Incentives Through Loss Aversion: A Field Experiment. *NBER working paper 18237*.

Fudenberg, D. and D. K. Levine (2006). A Dual-Self Model of Impulse Control. *American Economic Review 96*(5), 1449–1476.

Gill, D. and V. Prowse (2012, February). A Structural Analysis of Disappointment Aversion in a Real Effort Competition. *American Economic Review 102*(1), 469–503.

Golman, R. and G. Loewenstein (2012). Expectations and Aspirations: Explaining Ambitious Goal-setting and Nonconvex Preferences. *mimeo*.

Guiteras, R. P. and B. K. Jack (2014). Incentives, Selection and Productivity in Labor Markets: Evidence from Rural Malawi. *NBER Working Paper 19825*.

Hannan, R. L., V. B. Hoffman, and D. V. Moser (2005). Bonus Versus Penalty: Does Contract Frame Affect Employee Effort? In A. Rapoport and R. Zwick (Eds.), *Experimental Business Research, Vol II*, Volume II, Chapter 8, pp. 151–169. Netherlands: Springer.

Herweg, F., D. Müller, and P. Weinschenk (2010, December). Binary Payment Schemes: Moral Hazard and Loss Aversion. *American Economic Review 100*(5), 2451–2477.

Hilken, K., K. D. Jaegher, and M. Jegers (2013). Strategic Framing in Contracts. *Tjalling C. Koopmans Research Institute Discussion Paper Series 13-04*.

Horton, J. J., D. G. Rand, and R. J. Zeckhauser (2011, February). The online laboratory: conducting experiments in a real labor market. *Experimental Economics 14*(3), 399–425.

Hossain, T. and J. A. List (2012, July). The Behavioralist Visits the Factory: Increasing Productivity Using Simple Framing Manipulations. *Management Science 1909*, 1–17.

Hossain, T. and J. Morgan (2006). ...Plus Shipping and Handling: Revenue (Non) Equivalence in Field Experiments on eBay. *Advances in Economic Analysis & Policy 6*(2), Article 3.

Jayaraman, R., D. Ray, and F. de Vericourt (2014). Productivity Response to a Contract Change. *NBER working paper 19849*.

Just, D. and S. Wu (2005). Loss aversion and reference points in contracts. *mimeo*.

Kahneman, D. and A. Tversky (1979, January). Prospect Theory: An Analysis of Decision under Risk. *Econometrica 47*(2), 263–292.

Karlan, D. S. and J. Zinman (2009). Observing Unobservables: Identifying Information Asymmetries With a Consumer Credit Field Experiment. *Econometrica 77*(6), 1993–2008.

Kaur, S., M. Kremer, and S. Mullainathan (2013). Self-Control at Work. *mimeo*.

Kőszegi, B. (2014). Behavioral Contract Theory. *Journal of Economic Literature Forthcoming*.

Kőszegi, B. and M. Rabin (2006, November). A Model of Reference-Dependent Preferences. *The Quarterly Journal of Economics 121*(4), 1133–1165.

Kőszegi, B. and M. Rabin (2007, September). Reference-Dependent Risk Attitudes. *American Economic Review 97*(4), 1047–1073.

Koch, A. K., J. Nafziger, A. Suvorov, and J. van de Ven (2012). Self-Rewards and Personal Motivation. *Working paper*.

Lazear, E. (2000, February). Performance pay and productivity. *The American Economic Review 90*(5), 1346–1361.

Lazear, E. P. (1995). *Personnel Economics*. MIT Press.

Loewenstein, G. and D. Adler (1995, July). A Bias in the Prediction of Tastes. *The Economic Journal 105*(431), 929.

Loewenstein, G., T. O'Donoghue, and M. Rabin (2003, November). Projection Bias in Predicting Future Utility. *The Quarterly Journal of Economics 118*(4), 1209–1248.

Luft, J. (1994, September). Bonus and penalty incentives contract choice by employees. *Journal of Accounting and Economics 18*(2), 181–206.

Pope, D. and M. Schweitzer (2011). Is Tiger Woods loss averse? Persistent bias in the face of experience, competition, and high stakes. *The American Economic Review 101*(February), 129–157.

Rabin, M. (2000). Risk Aversion and Expected-Utility Theory: A Calibration Theorem. *Econometrica 68*(5), 1281–1292.

Salant, Y. and R. Siegel (2013). Contracts with Framing. *Working paper*.

Thaler, R. H. and H. Shefrin (1981). An Economic Theory of Self-Control. *Journal of Political Economy 89*(2), 392–406.

Figures

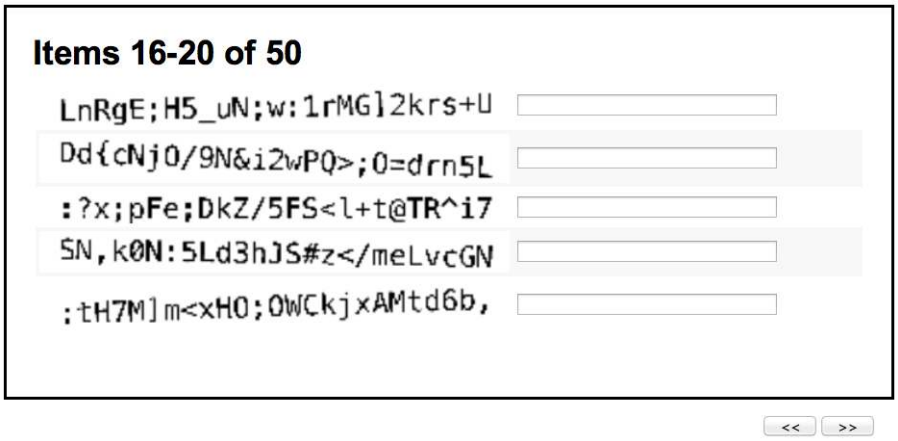
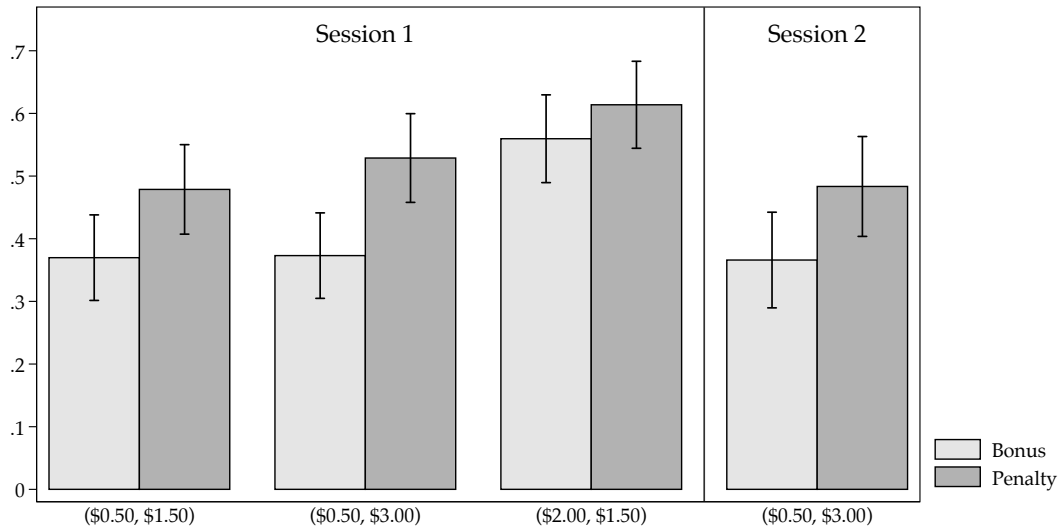


Figure 2: Example screen from the typing task.



Notes: financial incentive levels given in parentheses as (fixed pay, variable pay). Error bars indicate 95% confidence intervals.

Figure 3: Acceptance Rates by treatment.

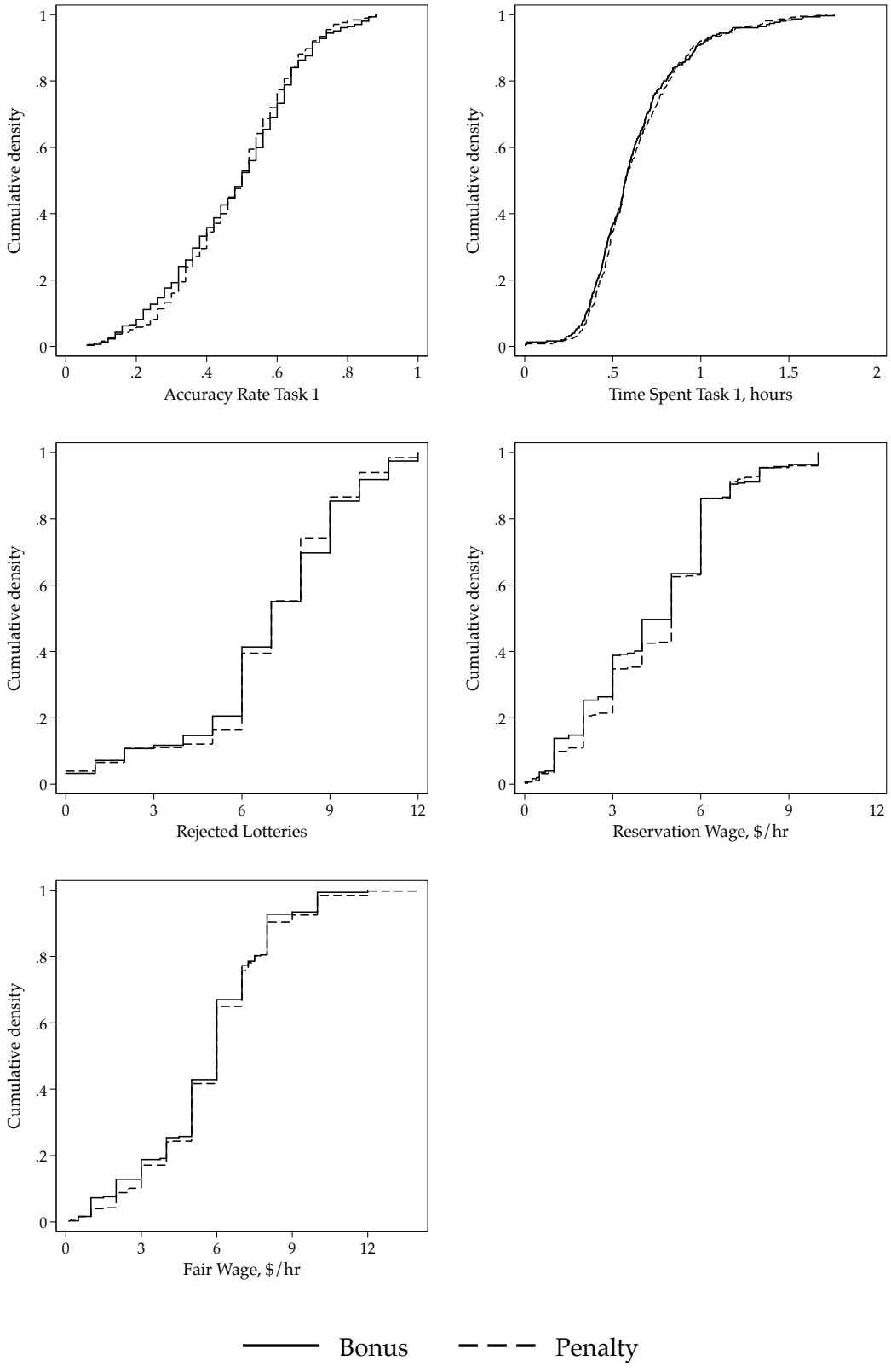
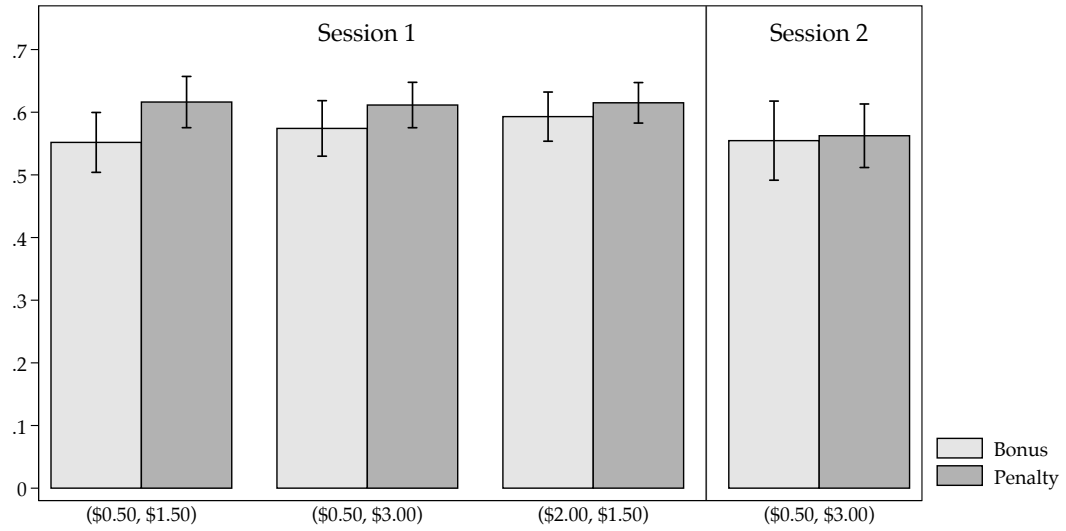


Figure 4: Comparing acceptors under Bonus and Penalty Frame. Reservation and fair wage trimmed at the 99th percentile.



Notes: financial incentive levels given in parentheses as (fixed pay, variable pay). Error bars indicate 95% confidence intervals.

Figure 5: Accuracy by treatment.

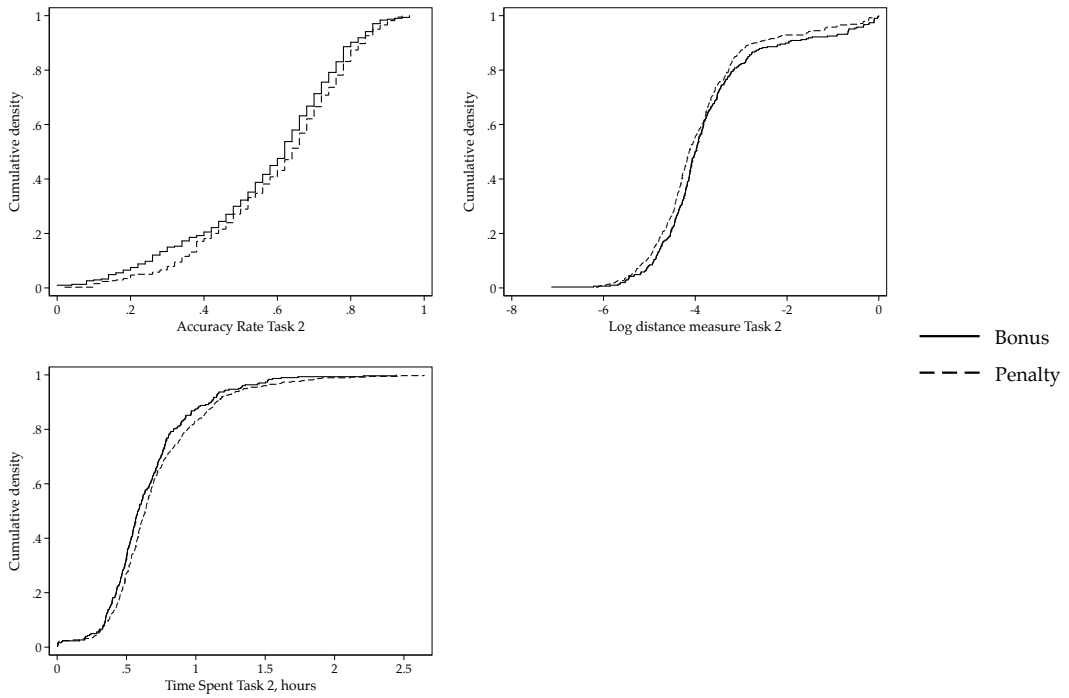
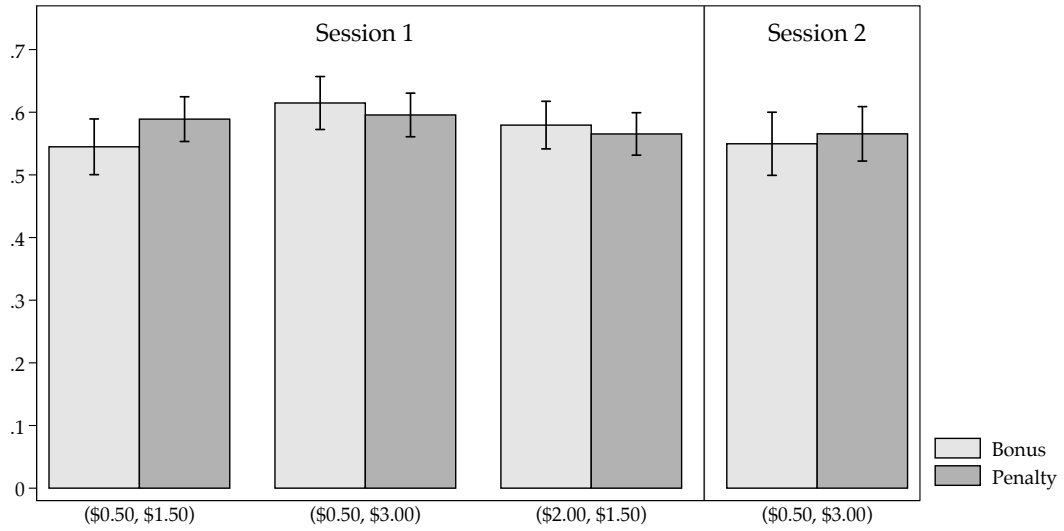
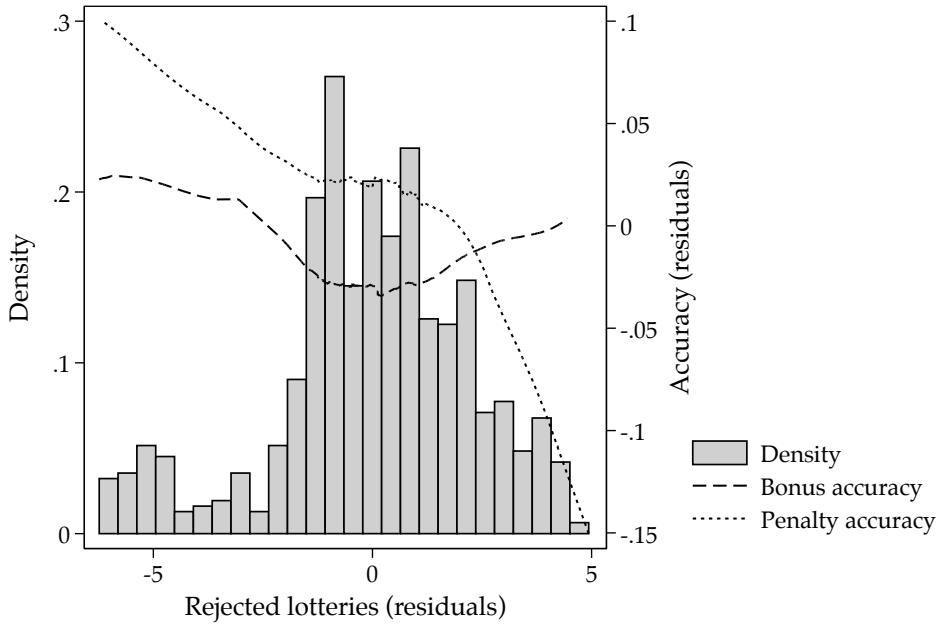


Figure 6: Performance and effort measures, comparing bonus vs penalty frame. Time spent is trimmed at the 99th percentile.



Notes: financial incentive levels given in parentheses as (fixed pay, variable pay). Error bars indicate 95% confidence intervals. The true mean task 1 performance was 0.46.

Figure 7: Participants' predictions of average task 1 accuracy by treatment.



Notes: figure plots the residual of stage 2 task accuracy against residuals of rejected lotteries, after partialling out all other controls and treatment effects. Participants who rejected no lotteries or all lotteries, and participants who made inconsistent choices are dropped.

Figure 8: Relationship between Rejected Lotteries and performance.

Tables

Table 2: Performance on effort tasks

	N	Median	Mean	s.d.
Round 1				
Accuracy Task 1	1450	0.46	0.46	0.18
Errors per item Task 1	1450	0.031	0.044	0.063
Hours on Task 1	1448	0.60	0.67	0.31
Round 1, Rejectors only				
Accuracy Task 1	763	0.44	0.43	0.18
Errors per item Task 1	763	0.033	0.048	0.070
Hours on Task 1	761	0.62	0.69	0.32
Round 1, Acceptors only				
Accuracy Task 1	687	0.50	0.48	0.17
Errors per item Task 1	687	0.028	0.039	0.053
Hours on Task 1	687	0.57	0.64	0.28
Round 2				
Predicted Mean Round 1 Accuracy	686	0.60	0.58	0.19
Accuracy Task 2	687	0.62	0.59	0.20
Errors per item Task 2	687	0.017	0.066	0.17
Hours on Task 2	686	0.62	0.75	1.24

"Accuracy" is the fraction of effort task items a participant entered correctly. "Errors per item" is computed as the mean of the Levenshtein distance between entered and correct answers, scaled by string length. "Hours on Task X" is estimated by multiplying the median page time by 10 to account for outliers. "Round 1, Acceptors (Rejectors) only" gives the stage 1 scores of the participants who accepted (rejected) the job offer in stage 2. "Predicted Mean stage 1 Accuracy" is the participant's response to the question "Of the 50 items in the typing task you did before, how many do you think people entered correctly, on average?" Where timing data is missing (two observations in stage 1, one in stage 2) this is due to JavaScript errors. One participant did not report a predicted stage 1 accuracy.

Table 3: Summary Statistics from stage 1 Survey

	N	Mean	s.d.
Loss Aversion			
Rejected Lotteries	1450	6.82	2.78
Inconsistent Lottery Choices	1450	0.07	0.25
Reservation Wage			
Reservation wage, $/hr	1450	4.86	2.70
Minimum fair wage, $/hr	1450	6.00	2.56
MTurk Experience			
Hours working on MTurk per week	1450	16.5	15.3
Typical MTurk earnings, $100/week	1450	0.58	0.58
MTurk HITs completed	1450	7457	27548
Months of experience on MTurk	1450	11.7	13.6
Mainly participate in research HITs	1450	0.77	0.42
Work on MTurk mainly to earn money	1450	0.93	0.26
Demographics			
Age in 2013	1449	32.8	10.7
Male	1450	0.49	0.50
Household Income	1450	38362	27942
Zipcode cluster	1450	0.13	0.34
Employment Status			
Full time	1450	0.37	0.48
Part time	1450	0.14	0.35
Self employed	1450	0.10	0.30
Full time MTurk worker	1450	0.11	0.31
Unemployed	1450	0.12	0.33
Student	1450	0.11	0.31
Other	1450	0.05	0.21
Education			
Less than High School	1450	0.00	0.07
High School / GED	1450	0.11	0.31
Some College	1450	0.32	0.47
2-year College Degree	1450	0.12	0.33
4-year College Degree	1450	0.35	0.48
Masters Degree	1450	0.07	0.25
Doctoral/Professional Degree	1450	0.02	0.13

"Rejected lotteries" is the number (0-12) of 50-50 win-lose lotteries that participants report they would be unwilling to play. "Inconsistent Lottery Choices" indicates participants who indicated inconsistent preferences over lotteries (accepting a lottery dominated by one they rejected). "Minimum acceptable wage" is the minimum hourly wage at which participants are willing to work on MTurk. "Minimum fair wage" is the reported minimum fair hourly wage that requesters "should" pay on MTurk. "Mainly participate in research HITs" indicates participants who report mostly working on HITs posted by researchers. "Work on MTurk mainly to earn money" indicates participants main reason for working ("for fun", "something else" coded as zero). "Male" is a dummy, note that six transgender individuals are coded as zeros. "Household income" is calculated using midpoints of income bins ("0-$30,000", then $10,000 bins until "> $100,000"). Zipcode cluster is a dummy indicating at least one other participant in the same session reported the same zipcode. "Employment Status" and "Education" are sets of mutually exclusive dummy variables.

Table 4: Balance Check

	Joint		Groups 0 & 1		Groups 2 & 3		Groups 4 & 5		Groups 6 & 7	
	F-stat	p	Diff	p	Diff	p	Diff	p	Diff	p
Accuracy Task 1	0.90	0.51	-0.01	0.54	0.00	0.94	-0.00	0.84	0.01	0.59
Hours on Task 1	0.51	0.83	0.01	0.76	-0.00	1.00	0.01	0.68	-0.04	0.28
Rejected Lotteries	0.53	0.81	0.31	0.26	-0.08	0.79	-0.23	0.39	-0.17	0.60
Inconsistent Lottery Choices	0.62	0.74	0.01	0.60	0.02	0.42	-0.01	0.80	0.01	0.81
Reservation wage, $/hr	0.65	0.72	0.07	0.77	-0.02	0.93	0.10	0.77	0.06	0.84
Minimum fair wage, $/hr	2.87	0.01	0.08	0.73	0.05	0.88	-0.38	0.14	0.16	0.59
Hours working on MTurk per week	1.65	0.12	-0.28	0.83	-1.61	0.21	-1.47	0.29	0.80	0.75
Typical MTurk earnings, $100/week	0.35	0.93	-0.04	0.54	0.01	0.79	0.01	0.84	-0.00	1.00
MTurk HITs completed	1.85	0.07	-7499.66	0.09	3609.55	0.07	460.60	0.71	-1966.82	0.50
Months of experience on MTurk	0.67	0.70	1.04	0.49	0.50	0.68	-0.34	0.80	-1.55	0.37
Mainly participate in research HITs	0.44	0.87	-0.04	0.39	0.04	0.38	-0.00	0.99	0.02	0.73
Work on MTurk mainly to earn money	1.22	0.29	-0.03	0.31	-0.03	0.25	0.00	0.96	0.02	0.52
Age in 2013	0.73	0.65	1.38	0.18	-0.80	0.45	-0.90	0.42	0.53	0.69
Male	0.93	0.48	0.08	0.12	-0.00	0.93	0.01	0.92	-0.01	0.81

"Joint" reports the F-statistic and p-value from a joint test of the significance of the set of treatment dummies in explaining each relevant baseline variable. The remaining columns report the difference in means and p-value from the associated t-test between pairs of treatment groups, where pairs differ only in terms of the bonus/penalty frame.

Table 5: Acceptance decision

	OLS				Logit (marginal effects)	
	(1) Accepted	(2) Accepted	(3) Accepted	(4) Accepted	(5) Accepted	(6) Accepted
Penalty Frame	0.110***	0.110***	0.116***	0.104***	0.110***	0.103***
	(0.026)	(0.026)	(0.029)	(0.026)	(0.025)	(0.025)
High Fixed Pay	0.163***	0.160***	0.152***	0.154***	0.161***	0.153***
	(0.036)	(0.036)	(0.041)	(0.036)	(0.035)	(0.034)
High Variable Pay	0.027	0.025	0.008	0.020	0.016	0.022
	(0.037)	(0.036)	(0.040)	(0.036)	(0.032)	(0.035)
Accuracy Task 1		0.312***	0.136	0.337		0.282***
		(0.075)	(0.352)	(0.309)		(0.074)
Accuracy Task 1 ^2			0.106	-0.055		
			(0.380)	(0.337)		
Hours on Task 1		-0.139***	-0.197***	-0.164***		-0.170***
		(0.042)	(0.059)	(0.044)		(0.047)
Rejected Lotteries		0.005	-0.007	0.004		0.005
		(0.013)	(0.015)	(0.013)		(0.013)
Reservation wage		-0.021***	-0.013	-0.016**		-0.018**
		(0.005)	(0.009)	(0.005)		(0.006)
Fair wage		0.004	-0.006	0.005		0.007
		(0.006)	(0.009)	(0.006)		(0.006)
Session 2	-0.026	-0.011	-0.002	-0.012		-0.012
	(0.038)	(0.038)	(0.041)	(0.038)		(0.037)
Inconsistent Lottery Choices				-0.065		-0.068
				(0.050)		(0.051)
Set dummies	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	No	Yes	Yes	No	Yes
N	1450	1448	1146	1447	1450	1447
R-squared	0.034	0.066	0.119	0.114		
Pseudo R-squared					.024	.089
Mean dep. variable	0.474	0.474	0.480	0.475	0.474	0.475

Dependent variable is a dummy indicating whether the participant accepted the job offer in stage 2. Columns (1)-(4) report OLS linear probability regressions, columns (5) and (6) present average marginal effects logit equivalents to (1) and (4). Standard errors clustered at zipcode-session level in parentheses. + p<0.10, * p<0.05, ** p<0.01, *** p<0.001. Two participants are missing timing variable and one missing age variable. Column (3) drops participants who made inconsistent lottery choices or who are above the 99th percentile of reservation wage, fair wage or time spent on task 1, or who are from zipcodes with more than one respondent in that session. "Set dummies" indicate the set of strings participants typed in stage 1. "Controls" are Hours worked on MTurk per week, MTurk earnings per week, HITs completed, MTurk experience, "Mostly do research HITs", "Mainly MTurk to earn money", age, male, and income, occupation and education dummies. "Rejected lotteries" measured in standard deviations.

Table 6: Acceptance: Interactions

	(1) Accepted	(2) Accepted	(3) Accepted	(4) Accepted
Penalty Frame	0.104***	0.128***	0.072*	0.107*
	(0.026)	(0.030)	(0.036)	(0.050)
High Fixed Pay	0.154***	0.200***	0.155***	0.190***
	(0.036)	(0.045)	(0.036)	(0.050)
Penalty * High Fixed Pay		-0.092		-0.071
		(0.059)		(0.071)
High Variable Pay	0.020	0.020	-0.014	0.004
	(0.036)	(0.036)	(0.045)	(0.048)
Penalty * High Variable Pay			0.068	0.032
			(0.051)	(0.062)
Set dummies	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
N	1447	1447	1447	1447
R-squared	0.114	0.115	0.115	0.115
Mean dep. variable	0.475	0.475	0.475	0.475

Dependent variable is a dummy indicating whether the participant accepted the job offer in stage 2. Estimates from OLS linear probability model. Standard errors clustered at zipcode-session level in parentheses. + p<0.10, * p<0.05, ** p<0.01, *** p<0.001. "Controls" are the full set of regressors from Table 5 specification (4). "Set dummies" indicate which set of strings participants typed in stage 1.

Table 7: Selection

	(1) Accepted	(2) Accepted	(3) Accepted	(4) Accepted	(5) Accepted	(6) Accepted	(7) Accepted
Penalty Frame	0.104***	0.104***	0.104***	0.104***	0.104***	0.104***	0.104***
	(0.026)	(0.026)	(0.026)	(0.026)	(0.026)	(0.026)	(0.026)
Accuracy Task 1	0.288***	0.245*	0.289***	0.288***	0.288***	0.287***	0.239*
	(0.076)	(0.101)	(0.076)	(0.076)	(0.076)	(0.076)	(0.102)
Penalty * Acc. Task 1		0.086					0.098
		(0.145)					(0.148)
Hours on Task 1	-0.164***	-0.165***	-0.156**	-0.164***	-0.166***	-0.165***	-0.167**
	(0.044)	(0.044)	(0.058)	(0.044)	(0.044)	(0.044)	(0.058)
Penalty * Hours Task 1			-0.018				0.000
			(0.082)				(0.083)
Rejected Lotteries	0.004	0.004	0.004	0.008	0.004	0.004	0.008
	(0.013)	(0.013)	(0.013)	(0.018)	(0.013)	(0.013)	(0.018)
Penalty * Rej. Lotteries				-0.008			-0.007
				(0.025)			(0.025)
Reservation wage	-0.016**	-0.016**	-0.016**	-0.016**	-0.020**	-0.016**	-0.020**
	(0.005)	(0.005)	(0.005)	(0.005)	(0.007)	(0.005)	(0.007)
Penalty * Res. wage					0.010		0.011
					(0.010)		(0.011)
Fair wage	0.005	0.005	0.005	0.005	0.004	0.003	0.004
	(0.006)	(0.005)	(0.006)	(0.006)	(0.006)	(0.007)	(0.007)
Penalty * Fair Wage						0.006	-0.001
						(0.010)	(0.012)
Set dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	1447	1447	1447	1447	1447	1447	1447
R-squared	0.114	0.114	0.114	0.114	0.114	0.114	0.115
Mean dep. variable	0.475	0.475	0.475	0.475	0.475	0.475	0.475

Dependent variable is a dummy indicating whether the participant accepted the job offer in stage 2. Estimates from OLS linear probability model. Standard errors clustered at zipcode-session level in parentheses. + p<0.10, * p<0.05, ** p<0.01, *** p<0.001. "Controls" are the full set of regressors from Table 5 specification (4) excluding Accuracy Task 1 squared. "Set dummies" indicate which set of strings participants typed in stage 1. Interaction variables are demeaned to stabilize interaction coefficients. "Rejected lotteries" is measured in standard deviations.

Table 8: Performance on stage 2

	(1) Accuracy	(2) Accuracy	(3) Accuracy	(4) Accuracy
Penalty Frame	0.036*	0.035**	0.025+	0.036**
	(0.015)	(0.012)	(0.014)	(0.012)
High Fixed Pay	0.024	0.032+	0.039*	0.038*
	(0.020)	(0.017)	(0.019)	(0.017)
High Variable Pay	0.014	0.023	0.018	0.025
	(0.022)	(0.018)	(0.022)	(0.018)
Accuracy Task 1		0.715***	0.986***	0.996***
		(0.037)	(0.188)	(0.172)
Accuracy Task 1 ^2			-0.296	-0.308+
			(0.190)	(0.176)
Hours on Task 1		-0.029	-0.043	-0.031
		(0.023)	(0.031)	(0.024)
Rejected Lotteries		-0.012+	-0.017*	-0.011+
		(0.006)	(0.007)	(0.006)
Reservation wage		-0.005+	-0.008*	-0.005+
		(0.003)	(0.004)	(0.003)
Fair wage		-0.001	0.000	-0.000
		(0.002)	(0.004)	(0.002)
Session 2	-0.038	-0.019	0.002	-0.016
	(0.024)	(0.019)	(0.021)	(0.019)
Inconsistent Lottery Choices				-0.073*
				(0.029)
Set dummies	Yes	Yes	Yes	Yes
Controls	No	No	Yes	Yes
N	687	687	550	687
R-squared	0.059	0.442	0.470	0.476
Mean dep. variable	0.590	0.590	0.596	0.590

Dependent variable is accuracy in the stage 2 typing task, measured as the fraction of items entered correctly. Standard errors clustered at zipcode-session level in parentheses. + p<0.10, * p<0.05, ** p<0.01, *** p<0.001. Column (3) drops participants who made inconsistent lottery choices or who are above the 99th percentile of reservation wage, fair wage or time spent on task 1, or who are from zipcodes with more than one respondent in that session. "Set dummies" indicate which set of strings participants typed in stages 1 and 2. "Controls" are Hours worked on MTurk per week, MTurk earnings per week, HITs completed, MTurk experience, "Mostly do research HITs", "Mainly MTurk to earn money", age, male, and income, occupation and education dummies. "Rejected lotteries" is measured in standard deviations.

Table 9: Stage 2 Performance: Interactions

	(1) Accuracy	(2) Accuracy	(3) Accuracy	(4) Accuracy
Penalty Frame	0.036**	0.033*	0.044**	0.044+
	(0.012)	(0.015)	(0.016)	(0.027)
High Fixed Pay	0.038*	0.032	0.038*	0.039
	(0.017)	(0.022)	(0.017)	(0.026)
Penalty * High Fixed Pay		0.010		-0.001
		(0.026)		(0.034)
High Variable Pay	0.025	0.025	0.035	0.035
	(0.018)	(0.018)	(0.025)	(0.028)
Penalty * High Variable Pay			-0.017	-0.018
			(0.025)	(0.034)
Set dummies	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
N	687	687	687	687
R-squared	0.476	0.476	0.477	0.477
Mean dep. variable	0.590	0.590	0.590	0.590

Dependent variable is accuracy in the stage 2 typing task measured as the fraction of items entered correctly. Standard errors clustered at zipcode-session level in parentheses. + p<0.10, * p<0.05, ** p<0.01, *** p<0.001. "Controls" are the full set of regressors from Table 8 specification (4). "Set dummies" indicate which set of strings participants typed in stages 1 and 2.

Table 10: Stage 2 Performance: Heterogeneous effects

	(1) Accuracy	(2) Accuracy	(3) Accuracy	(4) Accuracy	(5) Accuracy	(6) Accuracy	(7) Accuracy
Penalty Frame	0.037**	0.037**	0.037**	0.037**	0.037**	0.037**	0.037**
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
Accuracy Task 1	0.708***	0.745***	0.709***	0.708***	0.708***	0.705***	0.741***
	(0.037)	(0.053)	(0.037)	(0.037)	(0.037)	(0.038)	(0.054)
Penalty * Acc. Task 1		-0.071					-0.082
		(0.075)					(0.076)
Hours on Task 1	-0.031	-0.031	-0.019	-0.030	-0.031	-0.033	-0.010
	(0.024)	(0.024)	(0.036)	(0.024)	(0.024)	(0.024)	(0.037)
Penalty * Hours Task 1			-0.024				-0.040
			(0.044)				(0.046)
Rejected Lotteries	-0.011+	-0.011+	-0.011+	-0.005	-0.011+	-0.011+	-0.004
	(0.006)	(0.006)	(0.006)	(0.010)	(0.006)	(0.006)	(0.010)
Penalty * Rej. Lotteries				-0.012			-0.012
				(0.013)			(0.013)
Reservation wage	-0.005+	-0.005+	-0.005	-0.005	-0.004	-0.005+	-0.001
	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.003)	(0.004)
Penalty * Res. wage					-0.002		-0.009+
					(0.005)		(0.005)
Fair wage	-0.000	-0.000	-0.000	-0.001	-0.000	-0.003	-0.004+
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)
Penalty * Fair Wage						0.006	0.010*
						(0.004)	(0.004)
Set dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	687	687	687	687	687	687	687
R-squared	0.473	0.474	0.474	0.474	0.474	0.475	0.479
Mean dep. variable	0.590	0.590	0.590	0.590	0.590	0.590	0.590

Dependent variable is accuracy in the stage 2 typing task measured as the fraction of items entered correctly. Standard errors clustered at zipcode-session level in parentheses. + p<0.10, * p<0.05, ** p<0.01, *** p<0.001. "Controls" are the full set of regressors from Table 8 specification (4) excluding Accuracy Task 1 squared. "Set dummies" indicate which set of strings participants typed in stage 1. Interaction variables are demeaned to stabilize interaction coefficients. "Rejected lotteries" is measured in standard deviations.

Table 11: Expected earnings by treatment

Frame	Fixed Pay	Variable Pay	Expected Pay 1	Expected Pay 2	Accepted
Bonus	\$0.50	\$1.50	\$1.33	\$1.27	0.37
Penalty	\$0.50	\$1.50	\$1.42	\$1.36	0.48
Bonus	\$0.50	\$3.00	\$2.20	\$2.13	0.37
Penalty	\$0.50	\$3.00	\$2.27	\$2.21	0.51
Bonus	\$2.00	\$1.50	\$2.89	\$2.85	0.56
Penalty	\$2.00	\$1.50	\$2.92	\$2.91	0.61

Expected Pay 1 is the mean of $w + b * e$ for those who accepted the job, where e is Task 2 accuracy. Expected Pay 2 is the mean of $w + b * \hat{e}$ for all participants, where $\hat{e}$ are fitted values.

Appendices

A Model extensions

In the basic model I assume no diminishing sensitivity (i.e. utility is linear except for a kink at the reference point), that the reference point is entirely determined by P's choice of frame, and that there is no loss aversion in effort choice. In this appendix I discuss each of these candidate extensions in turn.

A.1 Diminishing sensitivity

Suppose that A's gain-loss utility exhibits diminishing sensitivity, à la Kahneman and Tversky (1979). Specifically, it is concave in the gain domain and convex in the loss domain. I assume that gain-loss utility for consumption x and reference point r equals $\mu(x-r)$ if $x \geq r$ and $-\lambda\mu(x-r)$ if $x < r$, where $\mu(.)$ is a concave function. For simplicity I assume consumption utility is still linear. The utility function (1) becomes:

$$U(e,w,b,F) = w + eb - \left(\frac{e^2}{2\gamma} - \alpha e\right) + e\mu((1-F)b) - \lambda(1-e)\mu(Fb) \tag{8}$$

Notice first that A's utility is still decreasing in F, since a higher reference point makes her worse off in every state of the world. Therefore firms will still tend to shy away from penalties, as in Proposition 1.

As for the choice of effort, it is slightly more complex than before. I obtain:

$$\begin{aligned} e^*(w,b,F) &= \gamma[\alpha + b + \mu((1-F)b) + \lambda\mu(Fb)] \\ \frac{de^*}{dF} &= \gamma b[\lambda\mu'(Fb) - \mu'((1-F)b)]. \end{aligned}$$

It is no longer guaranteed that e^* is increasing in F. $\frac{de^*}{dF} > 0 \leftrightarrow \lambda > \frac{\mu'((1-F)b)}{\mu'(Fb)}$, which will fail in the neighborhood of $F = 1$ if μ is sufficiently steep at zero. Intuitively, there are opposing effects to increasing the reference point: it makes A feel more disappointed when unsuccessful, but less elated when successful. Diminishing sensitivity implies that eventually the latter may outweigh the former, reducing effort, and further discouraging the use of penalties. Note however that $\frac{d^2e^*}{dFd\lambda} > 0$ as before, i.e. penalties have a stronger effect on more loss-averse people.

A.2 Reference-dependent effort provision

Crawford and Meng (2011) argue that reference dependence in effort provision may be an important determinant of labor supply, and that taxi drivers exhibit behavior consistent with both income and hours targeting. To capture this in a simple way, assume that A has an effort target e^r, and is pleased when actual effort is below the target level, and displeased when above.

The key question in this context is what effect the choice of *monetary* frame has on the effort target. If, as in Crawford and Meng (2011), A simply has an exogenous effort target, then the main results go through essentially unchanged. If the penalty frame decreases A's effort target then penalty framing will reduce her utility by even more than before. Lastly if the frame increases her target, it can actually be the case that the penalty frame increases her utility. The reduction of the loss (increase in the gain) in gain-loss utility over effort can offset the increase in loss (decrease in gain) of gain-loss utility over consumption.

I choose not to push this mechanism as a possible explanation for the main finding that penalties are more popular than bonuses, simply because I feel there is no a priori reason to think that a penalty frame over payoff outcomes would increase or decrease the effort target, and in particular that this effect would dominate the effect of loss aversion over consumption, which is the domain specifically targeted by the intervention.[41]

I assume that gain-loss utility in effort enters linearly, weighted by a parameter η. Note that since effort is a non-stochastic choice, A will be either always above, always below or exactly at her effort target. I assume that gain-loss *dis*utility in effort is $\lambda\eta(e - e^r)$ when $e \geq e^r$ and $(e - e^r)$ when $e < e^r$. A's utility function is now:

$$\begin{aligned} U(e, w, b, F) = w + e[\alpha + b(2 + (\lambda - 1)F)] - \lambda F b - \frac{e^2}{2\gamma} \\ - \eta(e - e^r)(\lambda(1 - \mathbb{1}[e < e^r]) + \mathbb{1}[e < e^r]) \end{aligned} \tag{9}$$

Differentiating U with respect to F yields $\frac{\partial U}{\partial F} = eb(\lambda - 1) - \lambda b + \eta \frac{de^r}{dF}(\lambda(1 - \mathbb{1}[e < e^r]) + \mathbb{1}[e < e^r])$. Apart from the discontinuity at $e = e^r$, this expression is maximized at $e = 1 \geq e^r$, equal to $-b + \eta\lambda \frac{de^r}{dF}$. Overall the sign on the expression is ambiguous, particularly since I lack a theory of how e^r depends on b and F.

A's optimal effort choice is as follows:

$$e^*(w, b, F) = \begin{cases} e^H & e^H \leq e^r \\ e^r & e^L < e^r < e^H \\ e^L & e^r \leq e^L \end{cases}$$

where $e^H = \gamma[\alpha + b(2 + (\lambda - 1)F] - \eta]$, and $e^L = \gamma[\alpha + b(2 + (\lambda - 1)F] - \eta\lambda]$.

Notice first that both e^H and e^L are increasing in F at rate $\gamma b\lambda$. Therefore a) if e^r is increasing in F, then e^* is everywhere increasing in F; b) if e^r does not depend on F, then e^* is increasing in F for $e^H < e^r$ or $e^L > e^r$ and flat when $e^L < e^r < e^H$, and lastly c) if e^r is *decreasing* in F, then e^* is first increasing, then decreasing, then increasing again.

[41] Moreover, a reasonable model for the setting of effort targets would be KR, such that the target is expected effort. However since effort is non-stochastic, expected and actual effort will always coincide so the gain-loss over effort term drops out altogether.

A.3 Expectations-based reference point (KR)

In the basic model I assume that the reference point is entirely determined by P's frame. However, KR argue that in many contexts it makes sense to think of *expectations* as a natural reference point. In this section I outline how to incorporate this intuition into the model.

In KR, the reference point is the expected distribution of outcomes. For example, an agent holding a lottery ticket that pays \$50 with 50% probability has a stochastic reference point that equals \$50 with probability one half and \$0 with probability one half. The most natural specification of the reference point in work on moral hazard under loss aversion applies Kőszegi and Rabin (2007)'s concept of choice-acclimating personal equilibrium (CPE).[42] Under CPE, the agent's choices influence her reference point, and she takes account of this effect when making her decision, i.e. her reference point and choices are simultaneously determined. For instance, under CPE, an agent knows that buying a lottery ticket will increase her expected winnings, which will be incorporated into her reference point. This creates the possibility of experiencing a loss when she does not win, which might ultimately lead her not to purchase a ticket. In the moral hazard context, higher effort leads to a higher chance of receiving b, which increases the reference point, and the agent anticipates this when choosing her effort level.

I continue to ignore risk aversion in the conventional sense, as well as diminishing sensitivity. I continue to assume that equal weight is placed on consumption and gain-loss utility (i.e. KR's η parameter equals one), and I continue to assume no loss aversion over effort choice. Formally, for reference point r distributed according to $H(r|e,F)$ A's utility function becomes:

$$\begin{aligned} U(G|H,e,F) = \quad & \underbrace{w + eb - \frac{e^2}{2\gamma} + \alpha e}_{\text{Consumption and effort cost}} \\ & + \underbrace{e\int \mu(w+b-r)dH(r|e,F) + (1-e)\int \mu(w-r)dH(r|e,F)}_{\text{Gain-loss utility}}. \end{aligned} \tag{10}$$

Where as before $\mu(x-r)$ equals $x-r$ if $x \geq r$ and $\lambda(x-r)$ if $x < r$.

In KR, the reference point is determined by recently held expectations, and in particular there is no mechanism for it to be influenced by framing. To allow for framing effects in the simplest possible way, I assume the reference point takes the form of a weighted sum of the KR reference point and the frame supplied by the principal.

[42] This is the specification used by Gill and Prowse (2012) and Herweg et al. (2010).

Formally, the marginal distribution of r is:

$$h(x|e,F) = Pr(r = x|e,F) = \begin{cases} 1-e & x = \phi w + (1-\phi)(w+Fb) \\ e & x = \phi(w+b) + (1-\phi)(w+Fb) \end{cases} \tag{11}$$
$$F \in [0,1], \phi \in [0,1).$$

As before, $w + Fb$ is the "base pay" supplied as a reference point by the principal, and F corresponds to the fraction of the bonus b that is presented as part of the base pay. $F = 0$ is a pure bonus frame with base pay w; $F = 1$ is a pure penalty frame with base pay $w + b$. ϕ captures how susceptible the agent is to framing effects. $\phi = 1$ coincides with KR's model, in which framing has no effect on behavior. $\phi = 0$ corresponds to the basic model in the text. For intermediate values of ϕ, A's reference point lies in between w and $w + b$ and she experiences mixed emotions whether she receives or does not receive the bonus.

Incorporating the above assumptions, A's gain-loss utility sums over four states of the world. With probability e^2, her consumption is $w + b$ and her reference point is $\phi(w+b) + (1-\phi)(w+Fb) \leq w + b$, putting her in the gain domain. With probability $e(1-e)$ her consumption is $w + b$ and her reference point is $\phi w + (1-\phi)(w+Fb) \leq w + b$, again putting her in the gain domain. With probability $e(1-e)$ her consumption is w and her reference point is $\phi(w+b) + (1-\phi)(w+Fb) \geq w$, putting her in the loss domain, and with probability $(1-e)^2$ her consumption is w and her reference point is $\phi w + (1-\phi)(w+Fb) \geq w$, again putting her in the loss domain. Plugging these into (10) and simplifying, I can write her utility as:

$$\begin{aligned} U(e,w,b,F) = e[\alpha + b((2-\lambda\phi) + (\lambda-1)(1-\phi)F)] - e^2\left[\frac{1}{2\gamma} - (\lambda-1)\phi b\right] \\ + w - \lambda(1-\phi)Fb. \end{aligned} \tag{12}$$

A.3.1 Effort choice

Suppose A accepts the contract. I impose two parameter restrictions that are sufficient for the optimal equilibrium effort choice to be interior (i.e. lie in the interval $(0,1)$) and given by the first-order condition.

Assumption 1 *No dominance of gain-loss utility:* $\lambda < \frac{2}{\phi} \equiv \hat{\lambda}$.

Assumption 2 *No top-coding in effort:* $\gamma < \bar{\gamma} \equiv \frac{1}{\bar{\alpha} + v[1+\lambda-\phi]}$.

Under Assumptions 1 and 2, A's optimal effort choice is equal to:[43]

$$e^*(b,F) = \frac{\gamma[\alpha + b((2-\lambda\phi) + (\lambda-1)(1-\phi)F)]}{1-2\gamma(\lambda-1)\phi b} \tag{13}$$

Assumption 1 ensures that A is not so strongly loss averse that she is discouraged from exerting any effort at all. This can happen because her reference point is increasing in her effort choice, as higher effort increases her expected earnings. Intuitively, if λ is large, the increase in her disappointment when unsuccessful outweighs the higher expected earnings, such that she prefers to not to exert any effort. Assumption 2 simply ensures $e^* < 1$.

It is easy to see that the three predictions of the basic model go through: effort is increasing in F, more strongly so for more loss-averse agents. A's utility is decreasing in F (just apply the envelope theorem to (12)), so Proposition 1 will hold.

The key difference now is that it is possible for effort to be decreasing in λ, as I find in the experiment. Differentiating e^* with respect to λ yields:

$$\frac{de^*}{d\lambda} = \frac{\gamma b[(1-\phi)F - \phi(1-2\gamma(\alpha + b(2-\phi)))]}{(1-2\gamma(\lambda-1)\phi b)^2} \tag{14}$$

The sign of this expression is ambiguous. In particular, suppose $F = 0$. Then $\frac{de^*}{d\lambda}$ is negative if $\phi > 0$ and $1 > 2\gamma(\alpha + b(2-\phi))$.

The intuition for this finding is the feedback between effort and the reference point in the model. On the one hand, increasing effort reduces the probability of experiencing a loss. On the other hand, as A increases her effort, she also increases her reference point, and thus the size of the loss when unsuccessful. It may that the latter effect outweighs the former, and hence that effort can be decreasing in λ. Note, however, that this does not change the prediction that the positive effect of incentive framing on effort is stronger for more loss-averse agents.

[43] To see that Assumptions 1 and 2 imply that e^* is the unique, interior maximizer of (12), first note that the second order condition can be written as $2\gamma(\lambda-1)\phi b < 1$. Assuming this holds, we have $e^* > 0$ for all F, α and $b > 0$, provided $2 - \lambda\phi > 0$, which is Assumption 1. Second, $e^* < 1$ for all F, α provided $\gamma b[(2-\lambda\phi) + (\lambda-1)(1-\phi) + 2(\lambda-1)\phi] < 1$, which can be easily seen to imply the second order condition when Assumption 1 holds, and therefore sufficiency of the first-order condition. Last, note that non-negative profits imply $b \leq v$, so substituting v for b and simplifying, the expression reduces to Assumption 2, which is therefore sufficient for $e^* < 1$ in equilibrium. Assumption 1 is the analog of the assumption of the same name in Herweg et al. (2010), whose model corresponds to the case $\phi = 1$.

B Additional Results

B.1 Summary distributions

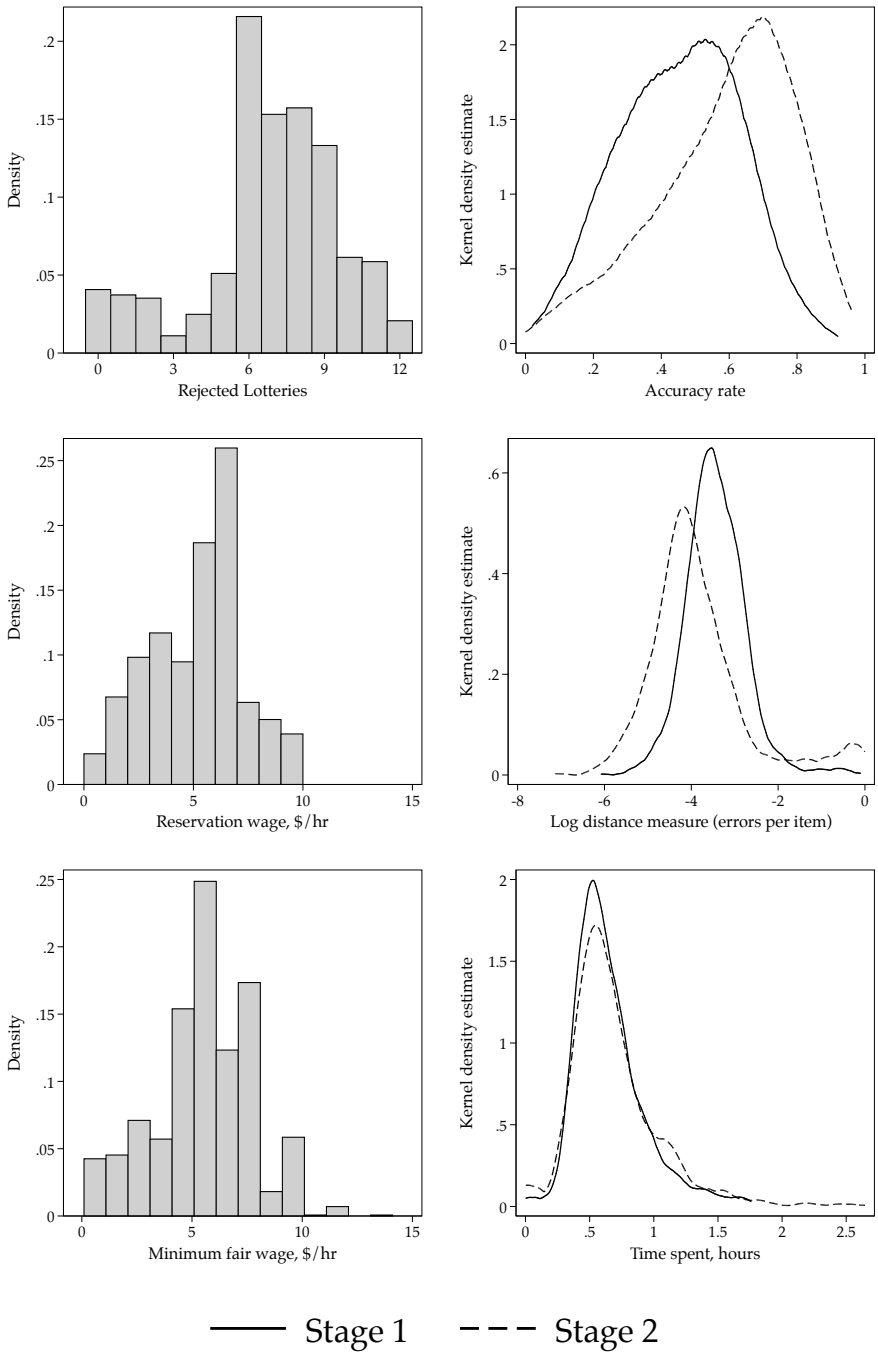


Figure B.1: Distributions of survey responses and task performance. Reservation wage, fair wage and time spent trimmed at the 99th percentile (minimum acceptable wage = \$10, minimum fair wage = \$14, time = 1.75 hours (round 1), 2.65 hours (round 2)).

B.2 Relation between stage 1 variables.

Before analyzing stage 2 behavior, it is instructive to analyze the relationships between the key variables measured in stage 1. The main variables of interest are the accuracy score, time spent on the effort task, the number of rejected lotteries, the reservation wage and the minimum fair wage reported by each participant. Table B.1 reports correlations between these variables with and without dropping participants who made inconsistent lottery choices. Higher performance on the stage 1 task is significantly negatively correlated with time spent on the task. Reservation wage and minimum fair wage are strongly positively correlated with one another. Reservation wage is negatively correlated with time spent on task 1, but not with performance. Lastly, there is no economically or statistically significant correlation between reservation wages, the number of rejected lotteries, and performance on the first stage task (rejected lotteries are significantly positively correlated with performance, but only when including participants who made inconsistent choices). These results suggest that selection on, say, loss aversion, will not necessarily imply selection on ability.

Table B.2 regresses stage 1 performance on various key variables. Time spent on stage 1 is robustly negatively correlated with performance on the effort task, with each additional hour spent associated with around 10 percentage points lower accuracy. Unlike the simple correlations, the minimum acceptable wage is now robustly negatively related with performance, perhaps because participants with higher reservation wages may be more likely to rush the task and therefore make more errors. A 1s.d. increase in the minimum acceptable wage ($2.70) is associated with 1.5-2.5 percentage points lower accuracy. In one specification, the minimum fair wage is positively associated with accuracy, controlling for reservation wages. Lastly, the number of rejected lotteries is positively associated with performance, although this is not statistically significant when participants who made inconsistent choices are dropped. The coefficient estimates imply a 0.6 to 1.3 percentage point improvement for a 1s.d. increase in the number of rejected lotteries.

Of the other variables, higher weekly Mturk earnings are negatively associated with performance, (although higher hours on MTurk positively associated, perhaps because those higher earnings are attained by expending less effort on each individual task). Participants who report that they mainly work on MTurk to earn money (93% of participants, other options “for fun”, “other”) perform around 5-6 percentage points better, men perform around four percentage points better, and participants who made inconsistent lottery choices around 8 percentage points worse.

Table B.1: Correlation between key stage 1 variables

Panel A

	Accuracy T1	Hours on T1	Rej. lotteries	Res. wage	Fair wage
Accuracy T1	1				
Hours on T1	-0.174***	1			
Rej. lotteries	0.0678**	-0.0275	1		
Res. wage	-0.0403	-0.107***	-0.0864***	1	
Fair wage	-0.00588	-0.0426	-0.0446	0.513***	1
Observations	1450				

Panel B

	Accuracy T1	Hours on T1	Rej. lotteries	Res. wage	Fair wage
Accuracy T1	1				
Hours on T1	-0.173***	1			
Rej. lotteries	0.0429	-0.0160	1		
Res. wage	-0.0348	-0.114***	-0.0786**	1	
Fair wage	0.00456	-0.0414	-0.0386	0.507***	1
Observations	1351				

Correlation matrices. * p<0.05, ** p<0.01, *** p<0.001. Accuracy T1 is accuracy on the stage 1 typing task. Hours on T1 is the median time spent on a page of the stage 1 task, multiplied by 10. Panel A includes all participants, Panel B drops participants who made inconsistent lottery choices.

Table B.2: Performance on stage 1

	(1) Accuracy	(2) Accuracy	(3) Accuracy	(4) Accuracy
Hours on Task 1	-0.107***	-0.131***	-0.096***	
	(0.015)	(0.022)	(0.018)	
Rejected Lotteries	0.009*	0.006	0.011*	0.013**
	(0.005)	(0.005)	(0.005)	(0.005)
Reservation wage	-0.005*	-0.009**	-0.005**	-0.005*
	(0.002)	(0.003)	(0.002)	(0.002)
Fair wage	0.002	0.006*	0.002	0.002
	(0.002)	(0.003)	(0.003)	(0.003)
Session 2	-0.020+	-0.017	-0.022+	-0.022+
	(0.011)	(0.013)	(0.012)	(0.012)
Hours working on MTurk per week			0.001	0.000
			(0.000)	(0.000)
Typical MTurk earnings, $100/week			-0.017+	-0.017+
			(0.009)	(0.009)
MTurk HITs completed			-0.000	-0.000
			(0.000)	(0.000)
Months of experience on MTurk			0.000	0.000
			(0.000)	(0.000)
Mainly participate in research HITs			0.002	-0.002
			(0.011)	(0.011)
Work on MTurk mainly to earn money			0.051**	0.057**
			(0.019)	(0.020)
Age			-0.000	-0.001
			(0.001)	(0.001)
Male			0.037***	0.038***
			(0.010)	(0.010)
Inconsistent Lottery Choices			-0.078***	-0.084***
			(0.017)	(0.017)
Constant	0.549***	0.562***	0.391***	0.337***
	(0.021)	(0.025)	(0.057)	(0.057)
Set dummies	Yes	Yes	Yes	Yes
Controls	No	No	Yes	Yes
N	1448	1147	1447	1449
R-squared	0.053	0.056	0.107	0.082
Mean of dependent variable	0.456	0.463	0.456	0.455

Dependent variable is accuracy rate on stage 1 typing task. Standard errors clustered at zipcode-session level in parentheses. + p<0.10, * p<0.05, ** p<0.01, *** p<0.001. All columns include dummies for the set of items typed. Column (2) drops participants who made inconsistent lottery choices, and those above the 99th percentile for time spent, reservation wage and fair wage. "Controls" are dummy variables for income, education and employment status. Rejected lotteries measured in standard deviations.

B.3 Balance checks

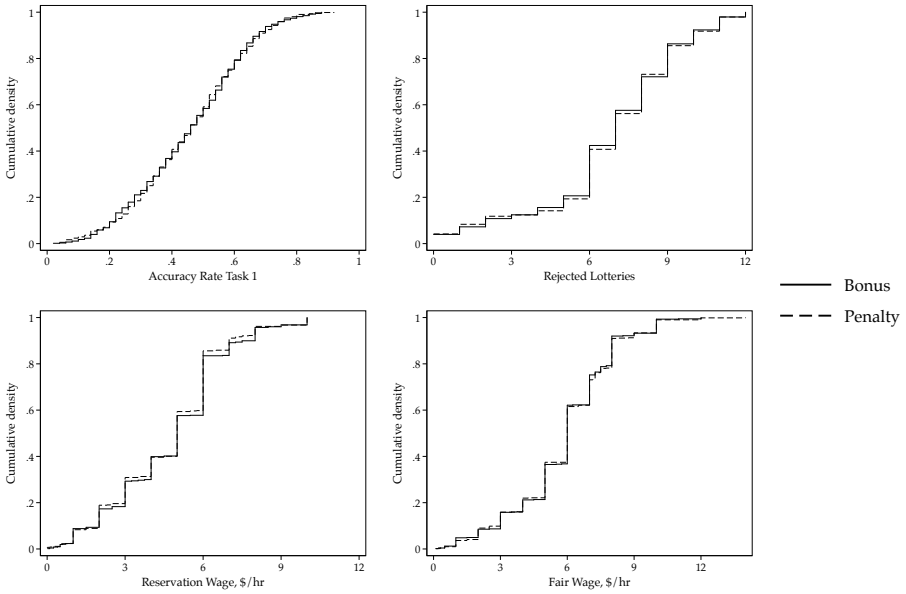


Figure B.2: Balance between bonus and penalty treatments, all incentive levels. Reservation wage (minimum acceptable wage) trims at the 99th percentile.

B.4 Comparing Session 1 and 2 Participants

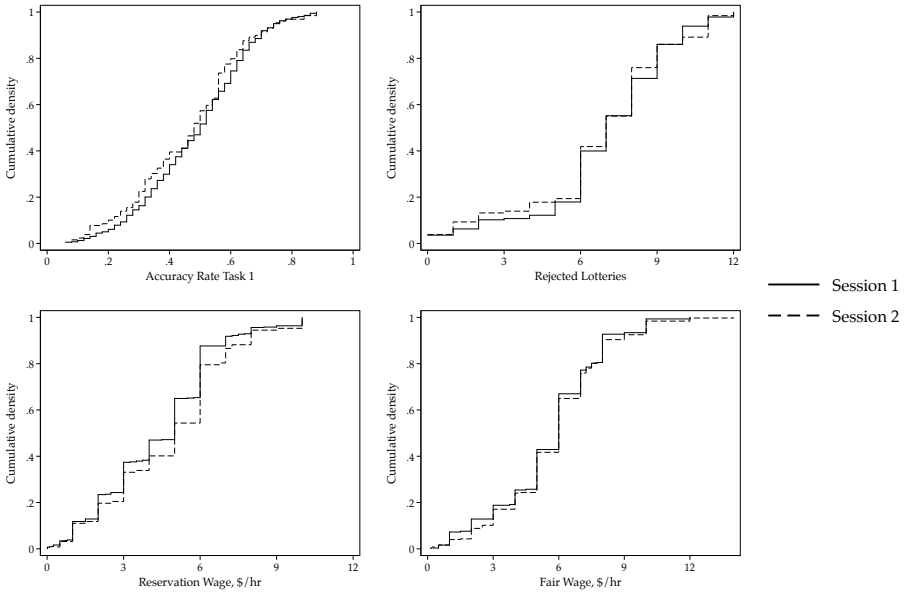


Figure B.3: Comparing session 1 and 2 participants. Reservation wage (minimum acceptable wage) trims at the 99th percentile.

B.5 Selection: Robustness

Table B.3: Selection: Robustness

	(1) Accepted	(2) Accepted	(3) Accepted	(4) Accepted
Penalty Frame	0.108***	0.102***	0.102***	0.105***
	(0.026)	(0.027)	(0.026)	(0.026)
Accuracy Task 1	0.286***	0.281***	0.278***	0.289***
	(0.076)	(0.078)	(0.076)	(0.076)
Hours on Task 1	-0.182**	-0.180***	-0.173***	-0.174***
	(0.069)	(0.046)	(0.043)	(0.044)
Penalty * Hours Task 1	0.018			
	(0.097)			
Rejected Lotteries	0.004	0.013	0.002	0.005
	(0.013)	(0.019)	(0.013)	(0.013)
Penalty * Rej. Lotteries		-0.021		
		(0.026)		
Reservation wage	-0.016**	-0.013**	-0.033***	-0.015**
	(0.005)	(0.005)	(0.009)	(0.005)
Penalty * Res. wage			0.021+	
			(0.012)	
Fair wage	0.005	0.003	0.004	-0.007
	(0.006)	(0.006)	(0.006)	(0.009)
Penalty * Fair Wage				0.014
				(0.012)
Set dummies	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
N	1433	1348	1433	1433
R-squared	0.113	0.114	0.117	0.115
Mean dep. variable	0.477	0.481	0.473	0.472

Dependent variable is a dummy indicating whether the participant accepted the job offer in stage 2. Estimates from OLS linear probability model. Standard errors clustered at zipcode-session level in parentheses. + p<0.10, * p<0.05, ** p<0.01, *** p<0.001. Column (1) drops participants above the 99th percentile of time spent on task 1. Column (2) drops those who made inconsistent lottery choices. Column (3) drops participants above the 99th percentile of reservation wage. Column (4) drops participants above the 99th percentile of fair wage. "Controls" are the full set of regressors from Table 5 specification (4) excluding Accuracy Task 1 squared. "Set dummies" indicate which set of strings participants typed in stage 1. Interaction variables are demeaned in each specification to stabilize interaction coefficients. "Rejected lotteries" is measured in standard deviations.

B.6 Alternative performance and effort measure

Table B.4: Performance/effort on stage 2, alternative measures

	(1) Log distance	(2) Log distance	(3) Time spent	(4) Time spent
Penalty Frame	-0.205*	-0.209**	0.054*	0.028
	(0.090)	(0.080)	(0.027)	(0.023)
High Fixed Pay	-0.251*	-0.323**	0.013	0.023
	(0.119)	(0.113)	(0.036)	(0.031)
High Variable Pay	-0.048	-0.069	0.026	0.036
	(0.133)	(0.130)	(0.040)	(0.033)
Accuracy Task 1				0.152
				(0.384)
Accuracy Task 1 ^2				-0.084
				(0.382)
Log Errors per item Task 1		0.868***		
		(0.066)		
Hours on Task 1				0.735***
				(0.067)
Rejected Lotteries		0.036		-0.018
		(0.040)		(0.012)
Reservation wage		0.025		0.001
		(0.021)		(0.006)
Fair wage		-0.010		-0.001
		(0.020)		(0.005)
Session 2	0.160	0.031	0.006	-0.035
	(0.149)	(0.127)	(0.046)	(0.036)
Inconsistent Lottery Choices		0.305		-0.036
		(0.189)		(0.042)
Set dummies	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes
N	687	687	680	680
R-squared	0.075	0.338	0.029	0.400
Mean dep. variable	-3.825	-3.825	0.687	0.687

Dependent variable in columns (1) and (2) is the log of the mean scaled Levenshtein distance (which can be interpreted as the mean number of errors per character in the typing task). Dependent variable in columns (3) and (4) is time spent on the stage 2 typing task, estimated as 10 times the median time spent per page of typed items, in addition dropping participants above the 99th percentile for time spent. Standard errors clustered at zipcode-session level in parentheses. + p<0.10, * p<0.05, ** p<0.01, *** p<0.001. "Set dummies" indicate which set of strings participants typed in stages 1 and 2. "Controls" are Hours worked on MTurk per week, MTurk earnings per week, HITs completed, MTurk experience, "Mostly do research HITs", "Mainly MTurk to earn money", age, male, and income, occupation and education dummies. "Rejected lotteries" is measured in standard deviations.

B.7 Participants' predictions of Task 1 accuracy

Table B.5: Participants' Predictions of stage 1 Accuracy

	(1) Predicted Acc.	(2) Predicted Acc.	(3) Accuracy	(4) Accuracy
Penalty Frame	0.002	0.001	0.036*	0.036**
	(0.015)	(0.014)	(0.015)	(0.012)
High Fixed Pay	0.006	0.010	0.022	0.037*
	(0.020)	(0.019)	(0.020)	(0.017)
High Variable Pay	0.033	0.036+	0.006	0.024
	(0.020)	(0.020)	(0.021)	(0.019)
Session 2	-0.040+	-0.035	-0.030	-0.017
	(0.022)	(0.022)	(0.024)	(0.019)
Predicted Acc.			0.201***	-0.006
			(0.040)	(0.032)
Set dummies	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes
N	686	686	686	686
R-squared	0.022	0.150	0.094	0.476
Mean of dependent variable	0.577	0.577	0.590	0.590

Dependent variable in (1) and (2) is participant's prediction of mean stage 1 accuracy, and in (3) and (4) is participant's actual stage 2 accuracy. Standard errors clustered at zipcode-session level in parentheses. + p<0.10, * p<0.05, ** p<0.01, *** p<0.001.
"Set dummies" indicate which set of strings participants typed in stage 1 (column (2)) or stages 1 and 2 (column (4)). "Controls" are the full set of regressors from the main specifications.

B.8 Comparing sessions 1 and 2

Table B.6: Comparing Sessions 1 and 2

	(1) Accepted	(2) Accepted	(3) Accuracy	(4) Accuracy	(5) Log distance	(6) Log distance	(7) Time spent	(8) Time spent
Penalty	0.156**	0.137**	0.037	0.039+	-0.317+	-0.238	0.051	0.040
	(0.052)	(0.052)	(0.029)	(0.023)	(0.186)	(0.168)	(0.055)	(0.049)
Session 2	-0.007	-0.015	-0.020	-0.000	-0.032	-0.062	-0.043	-0.061
	(0.053)	(0.053)	(0.039)	(0.030)	(0.242)	(0.198)	(0.057)	(0.052)
Penalty * Session 2	-0.038	0.006	-0.029	-0.028	0.330	0.149	0.089	0.045
	(0.076)	(0.075)	(0.051)	(0.038)	(0.308)	(0.256)	(0.087)	(0.071)
Set dummies	No	Yes	No	Yes	No	Yes	No	Yes
Controls	No	Yes	No	Yes	No	Yes	No	Yes
N	688	1447	302	687	302	687	300	680
R-squared	0.021	0.115	0.012	0.477	0.013	0.343	0.017	0.402
Mean dep. variable	0.439	0.475	0.580	0.590	-3.729	-3.825	0.698	0.687

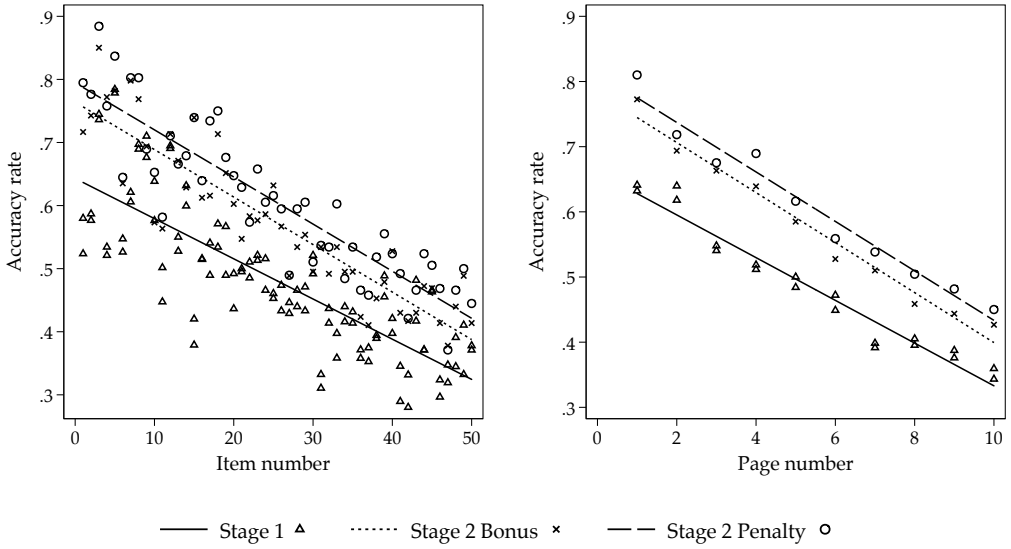
This table compares outcomes between treatment groups 2 and 3 (session 1) with groups 6 and 7 (session 2). Each received the same economic incentives (fixed pay of \$0.50 and variable pay of \$3.00), but the job offer was rephrased in session 2 to explore whether inattention might explain the session 1 results. Dependent variable in columns (1) and (2) is a dummy indicating whether the participant accepted the job offer in stage 2; in columns (3) and (4) is the participant's accuracy score in stage 2; in columns (5) and (6) is the "distance" accuracy measure described in table B.4 and in columns (7) and (8) is the time spent by the participant on the task in stage 2. Columns (7) and (8) drop participants above the 99th percentile of time spent. Standard errors clustered at zipcode-session level in parentheses. + p<0.10, * p<0.05, ** p<0.01, *** p<0.001. Odd-numbered columns do not include controls and use data from groups 2, 3, 6 and 7. Even-numbered columns include the full sets of controls from the equivalent main specifications, and use data from all participants to increase precision of estimated coefficients on control variables, differencing out the main effects for treatment groups 0, 1, 4 and 5 using dummy variables.

B.9 Does the framing effect wear off?

One interesting possibility discussed in Hossain and List (2012) is that the effect of an incentive frame might wear off over time as the agent realizes that the framing manipulation has no economic content.[44] Since my participants only perform the task once under framed incentives, I can only test for this within task, but nevertheless it is interesting to see whether performance under the bonus and penalty frame converges toward the end of the stage 2 task.

In Figure B.4 I plot performance by typed item, or by page of five typed items in stage 1, and separately for each framing treatment in stage 2, including only participants who accepted the stage 2 offer. The lines slope down because the text items grow progressively longer between pages. The graph clearly illustrates the shift in performance in the penalty over the bonus frame is consistent throughout the task, there is no evidence of convergence. I confirm this in a performance regression in Table B.7, where I find that the coefficient on item number interacted with the penalty dummy is a precisely estimated zero, while convergence would imply a negative coefficient.

Of course, this result does not imply that in the longer-run, for example after a couple of stages of incentive pay, participants' reference points would not shift to their true expected earnings, eliminating the framing effect. However over the short horizon of this experiment the penalty frame effect is persistent.



Notes: Plots the mean of performance for each typed item (page of 5 items) in stage 1, and each typed item (page) by framing treatment in stage 2.

Figure B.4: Performance by item/page on the effort task.

[44] See also Jayaraman et al. (2014) who find a short lived "behavioral" response to a contract change.

Table B.7: Performance by item in effort task

	(1) Item correct
Penalty Frame	0.038**
	(0.014)
Item	0.006***
	(0.002)
Item x Penalty	0.000
	(0.000)
Set dummies	Yes
Page dummies	Yes
Controls	Yes
N	34350
R-squared	0.126
Mean of dependent variable	0.590

Dependent variable is a dummy indicating whether an item was entered correctly. Standard errors clustered at zipcode-session level in parentheses. + p<0.10, * p<0.05, ** p<0.01, *** p<0.001. "Set dummies" indicate sets of strings participants typed in stage 1 and 2. "Page dummies" indicate the page of the current item. "Controls" are the full set of regressors from the main specifications, plus High Fixed and High Variable pay dummies.

B.10 Standard Selection Effect

In this section I briefly show that I do find a standard selection effect: that higher ability workers are more likely to accept the stage 2 offers of incentive pay, as also found by Lazear (2000) and Dohmen and Falk (2011), for example.

Figure B.1 shows the distribution of accuracy measures in stages 1 and 2, showing a large increase in performance in stage 2. This increase depends on three things: the effect of incentive pay on effort, the effect of incentive pay on selecting in motivated or able workers, and learning by doing. I cannot separate out learning by doing since I do not have a flat pay incentive treatment in stage 2, however I can illustrate the effect of selection.

Figure B.5 plots CDFs of stage 1 variables, comparing acceptors with rejectors (pooling all treatments). Acceptors performed better in stage 1, spent less time, and have lower reservation and fair wages, each one suggesting a first-order stochastic dominance shift. As previously noted I see little difference in rejected lotteries, which is surprising since the incentive pay is inherently risky. The differences in ability are also demonstrated by comparing stage 1 performance measures between acceptors and rejectors in Table 2

Ideally, to demonstrate the selection effect on performance I would regress accuracy on a "round 2" dummy, with and without controlling for ability. This is not possible since my main ability measure is performance on the stage 1 task, which is also an

outcome.[45] Instead, I perform the following simple exercise. I assume that the true performance model is linear and equal to that estimated in Table 8 column (4) (excluding the "set" dummy variables for the stage 2 task, since these are only observed for those who performed the task). Using this model I impute task 2 accuracy for rejectors.

The results are as follows. Mean accuracy across all participants in stage 1 is equal to 0.46, while in stage 2 it is equal to 0.59. Ignoring selection, a naive estimate of the combined effect of learning by doing and incentive pay would therefore be equal to a 13 percentage point improvement. However, the mean fitted stage 2 accuracy for all participants, including rejectors, is 0.56. Under the strong assumption that the fitted model is the true model, this implies that three percentage points of the combined effect can be attributed to advantageous selection of workers into incentive pay. An alternative way to control for selection is to compare mean performance of acceptors in stage 2 with mean performance of acceptors in stage 1, equal to 0.48. This gives me a similar effect of incentives and learning equal to 11 percentage points.

[45] Furthermore, I do not have a variable that will plausibly satisfy the exclusion restriction needed for a selection model.

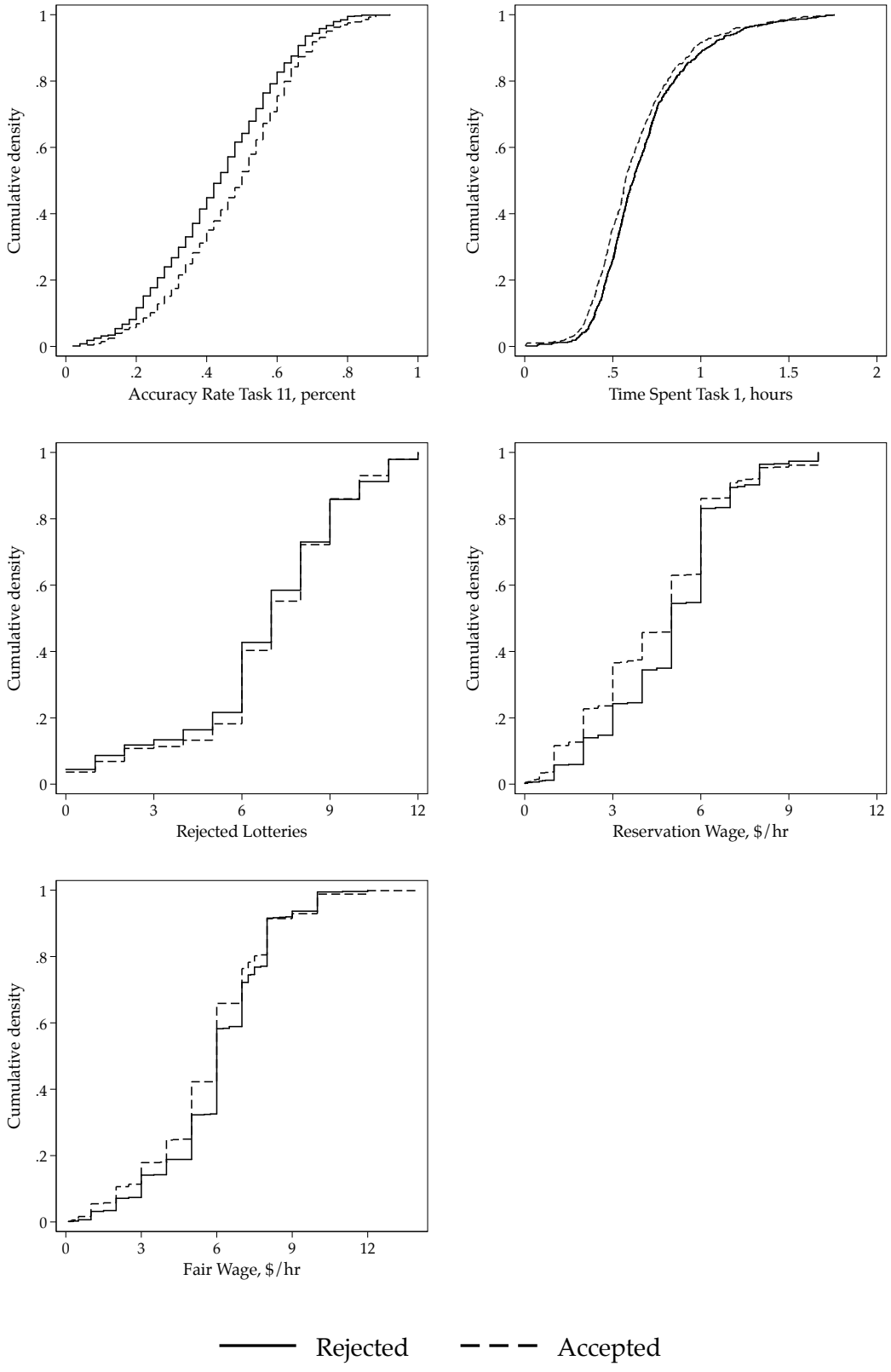


Figure B.5: Comparing Acceptors vs Rejectors. Reservation and fair wage trimmed at the 99th percentile.

B.11 Participant locations

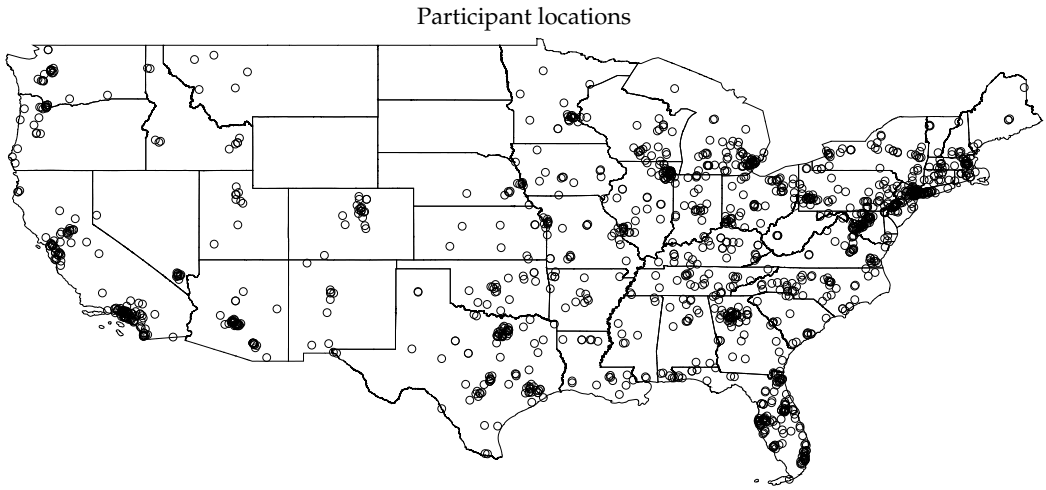


Figure B.6: Participant locations by geocoded zipcodes

B.12 Follow-up survey

Table B.8: Participants' perceptions of the job offer they received

Panel A

	The job offer or task is...							
	(1) Fun	(2) Easy	(3) Good pay	(4) Fair	(5) Good motivator	(6) Trustworthy	(7) Achievable	(8) Understandable
Penalty	-0.091	-0.104	0.295*	-0.171	-0.138	-0.045	-0.204+	-0.155
	(0.121)	(0.126)	(0.121)	(0.123)	(0.128)	(0.126)	(0.121)	(0.128)
Accepted offer	0.481***	0.366**	0.592***	0.658***	0.697***	0.523***	0.763***	0.241+
	(0.128)	(0.140)	(0.141)	(0.134)	(0.139)	(0.139)	(0.131)	(0.139)
N	252	252	252	252	252	252	252	252
R-squared	0.286	0.198	0.260	0.279	0.257	0.250	0.280	0.155

Panel B

	Offer attractive because...			Offer unattractive because...		
	(1) Good pay	(2) Elated if receive $3.50	(3) Encourages effort	(4) Risky	(5) Disappointed if receive $0.50	(6) Difficult
Penalty	0.304*	0.232+	-0.108	-0.078	0.170	0.089
	(0.124)	(0.126)	(0.128)	(0.115)	(0.116)	(0.114)
Accepted offer	0.627***	0.437**	0.699***	-0.882***	-0.778***	-0.924***
	(0.142)	(0.147)	(0.138)	(0.129)	(0.126)	(0.131)
N	252	252	252	252	252	252
R-squared	0.276	0.219	0.275	0.358	0.328	0.316

Panel C

	Estimated aggregate behavior	
	(1) Est. acceptance rate	(2) Est. success rate
Penalty	-0.006	-0.014
	(0.032)	(0.037)
Accepted offer	0.202***	0.118**
	(0.035)	(0.043)
N	252	252
R-squared	0.321	0.179
Mean dep. variable	0.560	0.446

Standard errors clustered at zipcode level in parentheses. + p<0.10, * p<0.05, ** p<0.01, *** p<0.001. Dependent variables in Panels A and B are standardized measures of agreement, measured on a 1-7 scale. Panel C asked participants to estimate the fraction of participants who received the same job offer as themselves and accepted, and the performance of acceptors. All regressions include the full set of controls from the main specifications.

C Experimental details

C.1 Rates of pay

One potential concern about the study is the low size of the incentives used: the maximum a participant could earn in stage 1 is $3 and the maximum in stage 2 is $3.50. If a participant spent 30 minutes on stage 1 this implies a $6 hourly wage, significantly lower than the typical rates of pay in experimental labs, for example. The size of the incentive was selected, based on trial runs of the task, to match Amazon's suggested rate of pay of $6; the concern was that paying more would be perceived as unusual or not comparable to other jobs on MTurk. More critically, to study selection effects it is critical that rates of pay are low enough that some participants do not want to accept the job. The 47% acceptance rate in stage 2 suggests that rates were pitched about right. Taking into account the level of the incentives, the relative sizes of the fixed and variable pay in stage 2 were selected to be reasonably high-powered to emphasize the role of the penalty component of the contract.

C.2 Accuracy report after first stage

Shortly after the first stage was completed, subjects were sent an email informing them of their performance. The purpose of this was to ensure that they understood the difficulty level of the task and had at least a sense of their ability. An example message is given below:

```
Thanks for doing the typing task + survey HIT.

We have now processed the data and approved your work.  We estimate
that out of the 50 items, you entered 31 (62%) without errors.

Best wishes

Jon
```

C.3 Invitation to second stage

C.3.1 Session 1

One week after the first stage, all participants from the first stage were sent a second email inviting them to the second stage task under their randomly assigned incentive.

```
Thanks for participating in our recent typing task and survey.  You are
invited you to do another typing task (typing 50 text strings) exactly
like the one you did before.  There is no survey this time.
```

```
Pay:
The basic pay for the task is $3.50.  We will then randomly select one
of the 50 items for checking.  If you entered it incorrectly, the pay
will be reduced by $3.00.

If you would like to perform this task, please use the following
personalized link which will take you straight to the task.

https://lse.qualtrics.com/SE/?SID=SV_a5HXEhTVyucdg1f&MID=XXXXXXXXXXXX

Your MTurk ID (XXXXXXXXXXXX) will be recorded automatically.  If you
don't want to do the task, you can just ignore this message.

The task will remain open for 4 days from the time of this message.
Payments will be made through the MTurk "bonus system" within 48 hours
of the task closing.

Best wishes

Jon de Quidt

PS: We'll select the line to be checked using a random number
generator.  If you attempt the task more than once, only the first
attempt will be counted.
```

Following the link in the email took them to the experimental task, in which the first page contained the same text.

C.3.2 Session 2

In session 2, the invitation was altered slightly to explore whether inattention might be driving the results. The concern was that inattentive participants might quickly read the first sentence of the "Pay" paragraph, which gives the high (low) base pay under the penalty (bonus) frame, then ignore the rest and click through to the task. Then the high acceptance rate under the penalty frame might be explained by inattentive participants reading only the high base pay. This would be a concern for external validity since for longer term job offers where the stakes are higher, workers presumably read their contracts carefully. The challenge in addressing this concern is to find a manipulation that encourages participants to read the invitation carefully without also changing their reference point, since then it would not be clear which was driving any change in outcome.

The invitation and introduction were kept identical to session 1 except for the "Pay"

paragraph which was rephrased as follows:

```
Pay:
The pay for this task depends on your typing accuracy.  We will
randomly select one item for checking, and if it was entered
incorrectly, the pay will be reduced below the base pay.  The base
pay is $3.50 which will be reduced by $3 if the checked item is
incorrect.
```

This phrasing first sets out the structure of the pay scheme, then gives all of the financial details in a single sentence, keeping the base pay and bonus/penalty amount as close to one another as possible.

C.4 Informed consent

After clicking the link on MTurk, but before beginning stage 1 of the experiment, participants were required to read and agree to an informed consent statement, which is reproduced below.

> The task you are going to complete forms part of a study of the behavior of workers on MTurk. Data on your answers in the following task will be collected and analyzed by researchers at the London School of Economics.
>
> Your participation is anonymous and no sensitive data will be collected. In addition, worker IDs will be deleted from any published data. Participation is voluntary and you can choose to stop at any time. There are no risks expected from your participation.
>
> We would like you to complete a typing task and a short survey. For an average typing speed this should take around 30 minutes to complete. At the end you will be given a completion code. Please copy and paste the code into the HIT on MTurk to be paid. The payment for completion of the HIT is $3.
>
> We may also contact you through MTurk to invite you to complete other HITs. This will not be affected by what you do in this HIT.
>
> If you have any questions or concerns at any time, please feel free to contact the researcher, Jon de Quidt, at <MTurk contact address>.
>
> If you are happy to proceed, please type "ACCEPT" into the box below and click through to the next page.

C.5 Lottery Questions

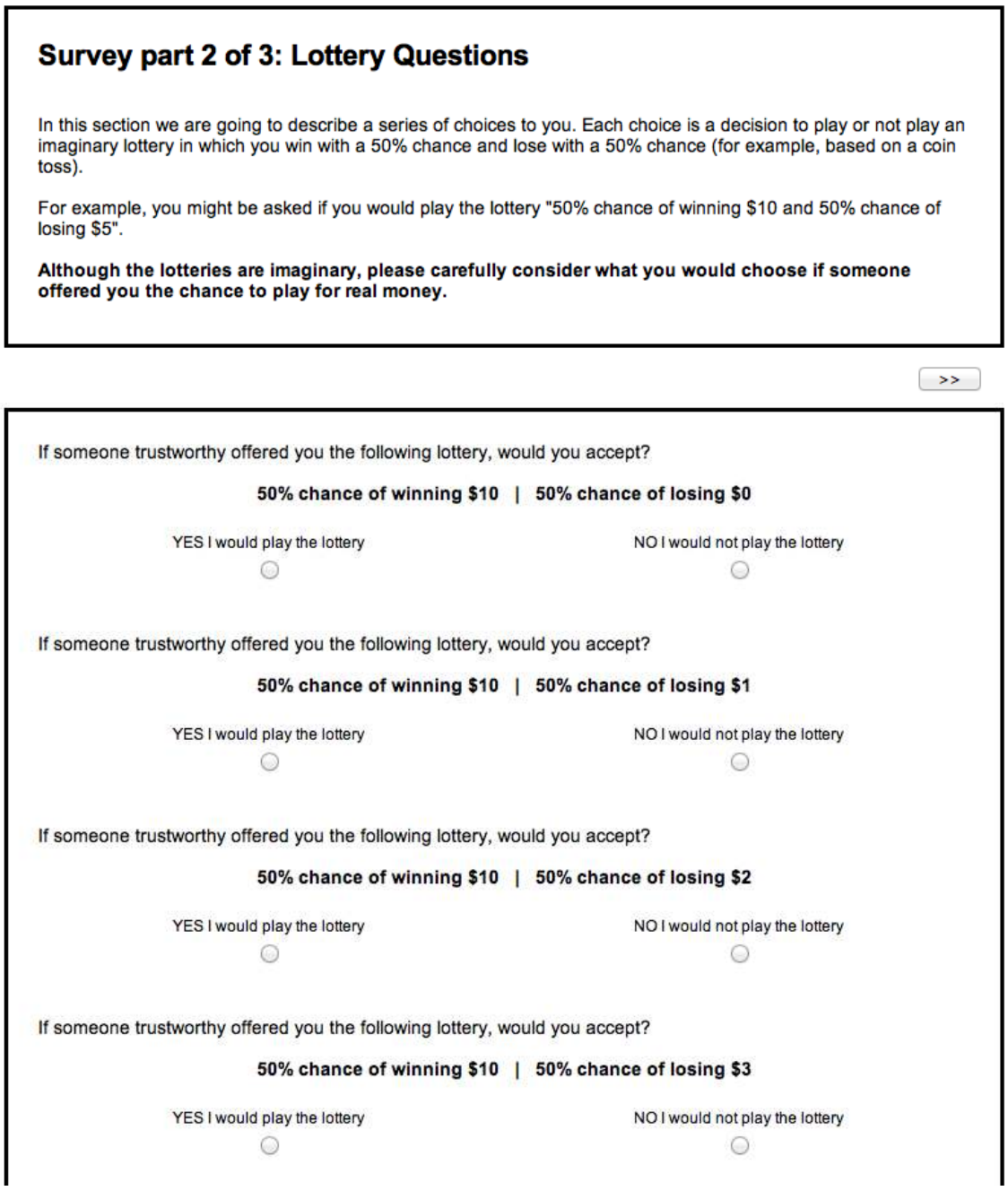


Figure C.1: Introduction and examples of lottery questions.