de Quidt, Jonathan

Working Paper

Your Loss is my Gain: A Recruitment Experiment with Framed Incentives

CESifo Working Paper, No. 6326

Provided in Cooperation with:
Ifo Institute – Leibniz Institute for Economic Research at the University of Munich

Suggested Citation: de Quidt, Jonathan (2017) : Your Loss is my Gain: A Recruitment Experiment with Framed Incentives, CESifo Working Paper, No. 6326, Center for Economic Studies and ifo Institute (CESifo), Munich

This Version is available at:
https://hdl.handle.net/10419/155568

The international platform of Ludwig-Maximilians University's Center for Economic Studies and the Ifo Institute

Working Papers

www.cesifo.org/wp

Your Loss is my Gain: A Recruitment Experiment with Framed Incentives

Jonathan de Quidt

CESIFO WORKING PAPER NO. 6326
CATEGORY 13: BEHAVIOURAL ECONOMICS
JANUARY 2017

ISSN 2364-1428

CESifo
Center for Economic Studies & Ifo Institute

CESifo *Working Paper No. 6326*

Your Loss is my Gain: A Recruitment Experiment with Framed Incentives

Abstract

As predicted by loss aversion, numerous studies find that penalties elicit greater effort than bonuses, even when the underlying payoffs are identical. However, loss aversion also predicts that workers will demand higher wages to accept penalty contracts. In six experiments I recruited workers online under framed incentive contracts to test the second prediction. None find evidence for the predicted distaste for penalty contracts. In four experiments penalty framing actually increased the job offer acceptance rate relative to bonus framing. I rule out a number of explanations, most notably self-commitment motives do not seem to explain the finding. Two experiments that manipulate salience are successful at eliminating the effect, but do not significantly reverse it. Overall, loss aversion seems to play surprisingly little role in this setting. The results also highlight the importance of behavioral biases for infrequent, binding decisions such as contract take-up.

JEL-Codes: D030, J410, D860.

Keywords: loss aversion, reference points, framing, selection, salience.

Jonathan de Quidt
Institute for International Economic Studies
Sweden - 106 91 Stockholm
jonathan.dequidt@iies.su.se

January 17, 2017
I acknowledge financial support from STICERD, Stiftelsen Lars Hiertas Minne, Stiftelsen Carl Mannerfelts fond and Handelsbanken's Research Foundations, grant no: B2014-0460:1. I thank many people for helpful discussions, particularly Philippe Aghion, Oriana Bandiera, Roland Bénabou, Tim Besley, Gharad Bryan, Tom Cunningham, Ernesto Dal Bó, Mathias Ekström, Florian Englmaier, Erik Eyster, Greg Fischer, Maitreesh Ghatak, Paul Heidhues, Dean Karlan, Matthew Levy, George Loewenstein, Rocco Macchiavello, Torsten Persson and Matthew Rabin, as well as many seminar and conference participants. All errors are my own.

1 Introduction

Loss aversion around a reference point (Kahneman and Tversky, 1979) is one of the best-known phenomena in behavioral economics. It has been employed to explain the willingness to accept/willingness to pay disparity or endowment effect (Kahneman et al., 1990); the equity premium puzzle (Benartzi and Thaler, 1995); small-stakes risk aversion (Rabin, 2000); labor supply behavior (Camerer et al., 1997) and much more. Reference-dependence implies that behavior depends not only on the mapping from actions into outcomes, but also on how those outcomes compare to a reference point. If reference points can be influenced by the way in which choices are described, or *framed*, then incentive design should consider the framing as well as the structure of incentives.

Consider two contracts for a task with known production function, the first of which 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 ignore the framing and behave identically under either one, but multiple studies find that workers exert higher effort under penalty contracts than equivalent bonus contracts.[1] Loss aversion neatly explains this behavior: losses loom larger than gains, so workers exert more effort to avoid a penalty than to achieve a bonus. The results hint at the possibility of costlessly nudging workers into higher productivity. Fryer et al. (2012) write that "there may be significant potential for exploiting loss aversion in the pursuit of both optimal public policy and the pursuit of profits."

These studies use already-recruited workers or lab subjects, so tell us only about incentive constraints. Without a meaningful participation decision (for example, nudging high school students to study more), this may be sufficient. But in the labor market, the worker's participation decision matters, determining the cost of recruitment and the types of workers recruited. I begin by showing that a simple reference-dependent model predicts workers will dislike penalty contracts, and demand higher wages to accept them. Intuitively, the high reference point induced by the penalty contract is costly because all outcomes under the contract become less satisfying: success is less rewarding and failure more painful.

This paper combines six experiments and subjective survey evidence to study preferences over framed contracts. In the main experiments I hire workers to perform a data entry task under gain- or loss-framed incentive contracts, and analyze how the framing affects the number and types of workers who are willing

[1] Church et al. (2008), Armantier and Boly (2015) and Hochman et al. (2014) in the lab, Hossain and List (2012) and Fryer et al. (2012) in the field.

to accept.

The bottom line is that I find no evidence of the predicted distaste for penalties. The first four experiments find *higher* take-up of penalty contracts. For example, workers offered a penalty contract for 30-40 minutes of typing were 25% more likely to accept than those offered an equivalent bonus contract. The final two experiments show that this effect can be eliminated by manipulating salience – a penalty contract naturally makes the higher base pay salient – either by changing the presentation of the contract terms or by asking workers to choose between a bonus and penalty contract presented side-by-side. In no experiment do I find a significant effect in the predicted, opposite direction.

Workers' behavior is otherwise conventional. Switching from flat pay to performance pay screened out low ability workers and improved typing accuracy by around 22%. Penalties elicited higher effort than bonuses—data entry accuracy was 6% or 0.2 s.d. higher under the penalty contract, a finding which is robust to controlling for selection. In fact, interestingly, I see no evidence of selection: bonus and penalty contract acceptors look essentially identical in a rich set of observables.

The sequence of experiments permits a detailed analysis of mechanisms. The first two robustness checks address rule out misunderstandings and differences in beliefs induced by the frame, combining experimental and survey evidence.

Next I turn to a explanation proposed in the literature for why workers might prefer penalties: commitment. Workers with weak self-control would like to exert high effort and penalties, by leveraging their loss aversion, help them to do so. Imas et al. (forthcoming) find that workers were willing to pay more for a penalty contract in which they could win or avoid losing a t-shirt, and argue that this is plausibly explained by a desire for commitment. Relatedly, field experiments by Kaur et al. (2010, 2015) find that a significant fraction of workers preferred a strictly dominated incentive contract that incorporated a commitment feature. However it is hard to explain my results with a desire for commitment. Notably, a coin-toss guessing task (experiment 3) in which performance is independent of effort and hence commitment cannot play a role, yet which replicates the higher take-up rate for the penalty contract.

Fourth, diminishing sensitivity in Prospect Theory generates risk seeking in the loss domain, as seen in the Tversky and Kahneman (1981) "Asian Disease" paradigm. If workers choose between a framed contract and safe outside option, potentially penalty framing increases the relative attractiveness of the contract. A contract valuation task (experiment 4) tests and does not support this mechanism.

Fifth I consider the role of salience.[2] Since the "base" or reference pay is higher under a penalty contract, it could be that workers are focusing on this outcome and underweighting possible deviations from it. Survey evidence suggests that workers subjectively perceive the penalty contract as "better paid," yet objectively understand that it is not. Two final experiments (experiments 5 and 6) seek to test the salience mechanism directly. Experiment 5 employs treatments that plausibly manipulate salience, by encouraging workers to focus on the contract as a whole and by de-emphasizing the base pay. Experiment 6 asks workers to directly choose between a bonus and penalty contract. Strikingly both experiments eliminate the penalty preference altogether. Experiment 5 also seems to eliminate the performance effect of penalties.

Across my six experiments and the parallel work of Imas et al. (forthcoming) (which I discuss in Section 6), incorporating variation in frame description, task, evaluation mode, worker experience and salience, *not one* found the theoretically predicted distaste for penalty contracts, more commonly finding the opposite effect. An important and relatively little-studied feature of reference-dependent preferences is how individuals make choices in anticipation of shifts in their reference point.[3] The literature so far reaches mixed conclusions. People seem not to anticipate how ownership will affect their reference point and willingness to trade (Loewenstein and Adler, 1995; Van Boven et al., 2000, 2003). On the other hand, they do seem to reduce effort provision if high effort is more likely to result in disappointment (Abeler et al., 2011; Gill and Prowse, 2012).[4] Here, I find little role for loss aversion in predicting contract take-up.

An additional finding of the paper is that penalty contracts can achieve higher performance at lower cost than bonus contracts – both take-up and performance were weakly higher in my experiments. It calls for more work to understand the costs of penalty contracts, given that they seem to be rarely used in practice (Baker et al., 1988; Lazear, 1991). One possible explanation is that the salience of specific outcomes or states of the world in real-world labor contracts is low.

[2]Note that "salience" here is distinct from the concept studied in the recent literature on context-dependent choice, e.g. Bordalo et al. (2012). It relates to existing work finding that, for example, people underweight non-salient sales taxes (Chetty et al., 2009) and eBay shipping costs (Hossain and Morgan, 2006). It also relates conceptually to the anchoring literature (e.g. Johnson and Schkade (1989), Ariely et al. (2003), Fudenberg et al. (2012), Mazar et al. (2013)) - possibly the high base pay of the penalty contract acts as a high anchor that increases willingness to accept. However, to my knowledge the existing literature focuses on anchors that are external to the object of interest and I am not aware of any work that tries to understand how attributes of the object itself can form an anchor.

[3]Most existing work studies behavior conditional on a reference point. Theories incorporating anticipation effects or "reference point management" include Karlsson et al. (2009), Köszegi and Rabin (2009), and Herweg et al. (2010).

[4]See also Ericson and Fuster (2011) and Heffetz and List (2014).

However, the contracts offered here were extremely transparent, consisting of just a task, an performance criterion and two possible payments. Real world contracts are more complex and it may be less transparent to workers why it is they have a "good feeling" about the contract under consideration.

It is reasonable to think that decision-making can become less biased with experience (e.g. List, 2004). But then we should be particularly interested in infrequent, contract acceptance or participation-type decisions, because of the potential for inexperienced people to make binding decisions that they later come to regret. While the contracts in this study are short-term, they highlight the normatively important possibility that people might be too willing to accept "exploitative" contracts. Indeed, after accounting for the time spent on the task, my penalty contract recipients earned *less* than bonus contract recipients. In other contexts, people might be over-willing to purchase durable goods or assets or accept free trials that then become difficult to part with (see e.g. Loewenstein et al., 2003). In a similar vein, Loewenstein et al. (2003) write that "people may be too prone to make reference-group-changing decisions that give them a sensation of status relative to their current reference group."

In what follows, Section 2 sets up a simple model of contracting with frames, and highlights three key testable predictions. Sections 3 and 4 present the experimental design and results. Section 5 analyzes mechanisms. Section 6 discusses external validity and related literature, then Section 7 concludes. Because of the large number of experiments the body of the paper contains only a restricted set of regression results and figures, I describe other results in the text and refer the interested reader to three Web Appendices containing additional theory, empirical results and experimental materials.

2 A simple model

The model is based on Herweg et al. (2010), who apply Kőszegi and Rabin (2006, 2007) (henceforth, KR) in a principal-agent setting.[5] In KR the reference point is the agent's rationally expected earnings distribution, which is invariant to framing. To incorporate framing while keeping the presentation simple, I assume for the main analysis that the agent's reference point is non-stochastic and a choice variable of the principal. The central prediction of a distaste for penalties relies only on penalty framing *increasing* the reference point relative to bonus framing.

An agent (A) is deciding whether to accept a contract to perform a task. If she

[5]For related analyses see the working papers by Just and Wu (2005) and Hilken et al. (2013).

rejects, she receives an outside option $\bar{u}$ which, for simplicity, requires no effort. If she accepts, she must exert an effort level $e \in [0,1]$, equalling the probability that the task is successful. If unsuccessful, the contract pays w, if successful it pays $w+b$.

The contract is *framed*, where the frame, F captures how the contract is described. The base pay is equal to $w+Fb$, the bonus if successful is $(1-F)b$ and the penalty if unsuccessful is Fb. Thus $F=0$ corresponds to a pure bonus frame where w is the base pay and b is a bonus for success. $F=1$ is a pure penalty frame where $w+b$ is the base pay and $-b$ is the penalty for failure. $F \in (0,1)$ is a mixed frame with both a bonus and penalty component.

A's utility is reference-dependent in money, evaluating monetary outcomes against a non-stochastic reference point, r. If she accepts the contract, her reference point is equal to the base pay, $r=w+Fb$. If she takes the outside option, $r=\bar{u}$. Her disutility of effort is not reference-dependent. The assumption that the reference point is that induced by the chosen option is the analog of KR's "choice acclimation;" in their theory, which does not admit framing effects, the reference point is the distribution of outcomes induced by the chosen option. In my context it amounts to assuming that the contract only induces sensations of gain or loss if it is accepted. As argued by KR and Herweg et al. (2010), choice acclimation is a natural assumption when payoffs are realized some time after choices (a few days later, in this experiment), so the agent knows her reference point will adapt to the choices she made.[6]

A's expected utility given r and a distribution of monetary outcomes y is:

$$U = E[y + G(y-r)|e] - c(e).$$

Utility consists of a standard component (expected earnings minus a convex cost of effort $c(e)$) and a gain-loss component, G, that evaluates monetary payoffs against the reference point. I assume $c(0)=0, c'>0, c''>0$. G has the basic properties described by Kahneman and Tversky (1979). It is "s-shaped": concave in the gain domain and convex in the loss domain, with a kink at zero:

$$G(x) = \begin{cases} \mu(x) & x \geq 0 \\ \lambda\mu(x) & x < 0 \end{cases}$$

where $\mu(0)=0$, $\mu'>0$. $\mu''(x) \leq 0$ for $x>0$ and $\mu''(x) \geq 0$ for $x<0$. In

[6] Bell (1985), Loomes and Sugden (1986) and Gul (1991) share the choice acclimation property. Web Appendix A allows the reference point to depend also on A's expected earnings (via a simple extension of KR) and discusses reference dependence in effort. Web Appendix A.3 shows the analog of prediction 3 using KR's Preferred Personal Equilibrium concept. Section B.8.1 discusses the implications of relaxing choice acclimation, and Experiment 4 provides a test.

words, utility is increasing in gains and decreasing in losses, but the marginal impact of gains and losses is (weakly) diminishing, a property referred to as diminishing sensitivity. λ captures the relative weight of losses to gains in utility. Loss aversion corresponds to $\lambda > 1$, whereby the disutility of a loss exceeds the utility of an equal-sized gain. I also make a simplifying symmetry assumption:

$$\mu(x) = -\mu(-x).$$

Finally, following KR and Herweg et al. I assume no probability weighting, so the expectation in U is taken with respect to the true distribution of y.

These assumptions imply very simple expressions for A's utility under the contract or outside option. Under the outside option, $U = \bar{u} + \mu(\bar{u} - \bar{u}) = \bar{u}$. Under the contract, A is successful with probability e and experiences a (weak) gain: $G(w + b - (w + Fb)) = \mu((1 - F)b)$. She is unsuccessful with probability $(1 - e)$ and experiences a (weak) loss: $G(w - (w + Fb)) = \lambda\mu(-Fb) = -\lambda\mu(Fb)$. Her utility can therefore be written as:

$$U(e, w, b, F) = w + eb - c(e) + e\mu((1 - F)b) - (1 - e)\lambda\mu(Fb). \tag{1}$$

Her optimal effort e^* solves the first order condition:[7]

$$b + \mu((1 - F)b) + \lambda\mu(Fb) - c'(e^*(b, F)) = 0. \tag{2}$$

A accepts a contract (w, b, F) if her participation constraint is satisfied:

$$U^*(w, b, F) - \bar{u} \geq 0 \tag{3}$$

Where $U^*(w, b, F) = U(e^*(b, F), w, b, F)$. For the nontrivial case where $b > 0$ this simple model yields three key testable predictions:

Prediction 1. *If A is loss averse ($\lambda > 1$) her effort is strictly higher under a pure penalty contract ($F = 1$) than a pure bonus contract ($F = 0$).*
Proof: $e^*(b, 1) - e^*(b, 0) = c'^{-1}(b + \lambda\mu(b)) - c'^{-1}(b + \mu(b)) > 0$.[8]

Prediction 2. *If $c(e)$ is quadratic, penalties have a more positive effect on effort*

[7] For simplicity, I assume b is low enough that the solution e^* is smaller than one.

[8] $\lambda > 1$ alone does not imply that e is monotone in F. To see this, observe that $\partial e^*(b, F)/\partial F = (\lambda\mu'(Fb) - \mu'((1 - F)b))/c''(e^*(b, F))$, which is positive if $\lambda\mu'(Fb) - \mu'((1 - F)b) > 0$. The numerator is decreasing in F, so if it is positive for $F = 1$, it is for all F, hence monotonicity is guaranteed if $\lambda > \mu'(0)/\mu'(b) \geq 1$. A sufficient condition is that μ is linear. Intuitively, diminishing sensitivity implies that outcomes far from the reference point are weighted less strongly than outcomes close to the reference point, so the incentives may be sharper with intermediate than extreme reference points. See also Armantier and Boly (2015).

This observation is important if we relax the assumption that the principal can freely choose A's reference point. Under the weaker assumption that he can manipulate r but only over some range, effort may not be higher under the penalty contract. Formally, suppose that a pure bonus frame now corresponds to F^B and a pure penalty frame to F^P, where $0 < F^B < F^P < 1$. It could then be that $e^*(b, F^P) < e^*(b, F^B)$.

for more loss-averse agents.
Proof: if c is quadratic, $\frac{\partial^2 e^*}{\partial F \partial \lambda} \propto b\mu'(Fb) > 0$.[9]

Prediction 3. *Penalty framing reduces A's willingness to accept the contract.*
Proof: by the envelope theorem, $\frac{dU^*}{dF} = -b[e^*\mu'((1-F)b)+(1-e^*)\lambda\mu'(Fb)] < 0$.[10]

Prediction 1 matches the findings in the existing literature on the effects of penalty framed incentives, and is also explored in the empirical analysis in this paper. The main focus of the paper is on Prediction 3. Agents may also be heterogenous, for example differing in loss aversion or the cost of effort, and the outside option may also depend on their type, so the empirical part of the paper studies whether types differentially select into penalty contracts. Since any increase in F reduces A's utility, prediction 3 does not rely on the principal being able to freely choose A's reference point, only that penalty framing increases it.

Interestingly, the following proposition shows that the participation effect dominates the incentive effect, such that it is more costly to elicit a given effort level using penalties than using bonuses.

Proposition 1. *Consider a contract* (w, b, F)*, where* $F > 0$*, that elicits effort level* e *and gives A utility* u*. Then, there exists an alternative contract,* (w', b', F')*, where* $F' < F$*, that elicits* e*, gives A at least* u *and where A's expected compensation is strictly lower, i.e.* $w' + eb' < w + eb$*. Therefore, the lowest-cost contract that elicits* e *is a pure bonus contract with* $F = 0$*.*

The proof is given in Web Appendix A. If Proposition 1 is correct, it could help to explain why firms appear reluctant to use penalty contracts.

3 Experimental design

The data come from six experiments with US-based workers on Amazon Mechanical Turk (MTurk, for short). MTurk is a large online labor market for "crowdsourcing."[11] A "requester" that needs data entered, audio recordings transcribed, images categorized, proofreading, et cetera, can post a job on MTurk and recruit "workers" to carry it out. Pay is set by the requester.

[9] For general cost functions $c'''(.)$ matters. $\partial^2 e^*/\partial F \partial \lambda = c''^{-1}\left(b\mu'(Fb) - c'''(e^*)\frac{de^*}{dF}\frac{de^*}{d\lambda}\right)$. Quadratic c implies $c''' = 0$.

[10] Note that this condition does not depend on de^*/dF and holds for all $\lambda \geq 0$. It relies only on reference dependence, not loss aversion, i.e. A must care about gains and losses but need not necessarily overweight losses. Intuitively: increasing the reference point decreases the utility of *all* outcomes–gains become less rewarding and losses more painful–irrespective of the relative weight applied to losses. Incorporating probability weighting would not change the result, as an increase in F decreases gain-loss utility however its components are weighted.

[11] Typically there are several hundred thousand tasks available for workers to perform. In 2011 Amazon reported that there were around 500,000 registered worker accounts.

MTurk enables testing for selection effects in a natural setting, where the worker has access to many alternative tasks as an outside option. Most work on MTurk is performed for low wages (my workers reported a mean reservation wage of $5.14 per hour, and mean typical hourly earnings of $5.89), enabling me to recruit a large sample. The average worker in my sample works for 17 hours per week on MTurk and has been a worker for 13 months (detailed summary statistics are given in Web Appendix Table B2). In general, MTurk workers have been found to be quite representative of the US population (Berinsky et al., 2012).

MTurk is increasingly commonly used for research by social scientists. To cite a couple of examples, Bordalo et al. (2012) test their theory of salience using MTurk; Horton et al. (2011), Amir et al. (2012) and Berinsky et al. (2012) replicate several classic experimental results on MTurk. Kuziemko et al. (2015) study preferences for redistribution and DellaVigna and Pope (2016) study a wide range of effort incentives, including gain/loss framing (they find a positive but non-significant effect of loss framing on effort).

3.1 Design specifics, experiments 1, 2 and 3

This section discusses experiments 1, 2 and 3, which share a common two-stage design, similar to Dohmen and Falk (2011). Section 4 then describes the results of those experiments. Experiments 4, 5 and 6 target specific mechanisms, so I describe them separately in Section 5. Prior participants are always blocked from participating in subsequent experiments.

Each experiment consisted of a first stage where workers were recruited on MTurk for a real-effort task and survey, and paid a fixed amount. The next day, they were sent their accuracy score by email. Then, a week later, workers from the first stage were sent a surprise job offer to perform the task again (stage 2), this time under framed performance pay. Workers were allowed four days to complete stage 2, were free to ignore the offer if not interested, and told that payments would be made within 48 hours of the four day window closing. Instructions and other experimental materials are reproduced in Web Appendix C. The design is summarized in Web Appendix Figure C1.

The task in experiments 1 and 2 was transcribing 50 text strings, increasing in length from 10 to 55 characters (example given in Web Appendix figure C2). The strings were generated using random combinations of letters, numbers and punctuation and distorted to give the appearance of having been scanned or photocopied.[12] The task was designed to take around 30 minutes of focused

[12]The task mimics CAPTCHA puzzles (Completely Automated Public Turing test to tell

effort, to be implementable online and be sufficiently difficult to avoid ceiling effects. There were 10 possible sets of strings in each stage, assigned to workers at random. The task in experiment 3 was guessing 50 coin tosses, and took around 10 minutes. Workers were not put under time pressure, nor was speed rewarded, to avoid multitasking concerns between speed and accuracy.

The stage 1 jobs posted on MTurk were advertised as a \$3 "typing task and survey" (experiments 1 and 2) or \$1 "guessing task and survey" (experiment 3). Stage 2 job offers consisted of a fixed pay component that did not depend on performance, a variable pay component paid if the accuracy check was passed, and a frame (bonus or penalty). Experiment 1 randomized the levels of fixed and variable pay to study how behavior responds to these terms, see Table 1 below.

Workers were told that after completion of the stage 2 task I would select, using a random number generator, one of the 50 strings or coin tosses that they had been assigned to type or guess. They would receive the bonus (avoid the penalty) conditional on that item being entered correctly. Hence the probability of receiving the bonus equals the accuracy rate. Only the "pay" text in the invitation differed between treatments. The key phrasing is given in Table 2, full email text in Web Appendix C.5. I deliberately avoided emotive words like "bonus" and "penalty."

Note that while low relative to most lab experiments, the pay rates were comparable to typical rates on MTurk, which is necessary to study participation decisions (if pay were too high, everyone would accept). Incentives were deliberately high powered to maximize power, and are detailed in Table 1.

The two stage design ensures that workers know the task and their ability, to control for inference about the nature of the task from the incentive contract they are offered. It ensures that workers have interacted with the principal (me) before, in case penalty offers are perceived as more or less trustworthy. It also enables me to measure types at baseline, to measure selection effects and to stratify the randomization for balance.

Experiments 2 and 3 collected additional data. In experiment 2, four days after stage 2 ended, workers were invited to a paid follow-up survey, to gain qualitative insight into possible mechanisms. In experiment 3, workers were invited to a third stage one week after stage 2, under the same terms and framing as stage 2, to study how framing effects persist.

Computers and Humans Apart), used to prevent bots from accessing websites.

Table 1: Treatments

Experiment	Group	N	Fixed pay	Variable pay	Frame
Experiment 1	0	192	\$0.50	\$1.50	Bonus
Data entry	1	188	\$0.50	\$1.50	Penalty
	2	193	\$0.50	\$3	Bonus
	3	191	\$0.50	\$3	Penalty
	4	193	\$2	\$1.50	Bonus
	5	189	\$2	\$1.50	Penalty
Experiment 2	6	153	\$0.50	\$3	Bonus
Data entry	7	151	\$0.50	\$3	Penalty
Experiment 3	8	202	\$0.30	\$1	Bonus
Coin toss	9	196	\$0.30	\$1	Penalty
Experiment 4	10	96	\$2*	\$2*	Bonus
Contract valuations	11	110	\$2*	\$2*	Penalty
Experiment 5	12	191	\$1	\$1	Bonus, conventional
Reduced salience	13	194	\$1	\$1	Penalty, conventional
	14	194	\$1	\$1	Bonus, table
	15	199	\$1	\$1	Penalty, table
Experiment 6	16	149	\$0.40	\$0.40	Both, weak preference
Direct choice	17	146	\$0.40	\$0.40	Both, strict preference

* Note: hypothetical payoffs.

3.2 Data collected

Summary statistics are presented in Web Appendix Tables B1 and B2.

I attempt to measure loss aversion with an unincentivized variant of the measure used by Gächter et al. (2010) and Abeler et al. (2011).[13] Workers indicated whether they would play each of 12 lotteries of the form "50% chance of winning \$10, 50% chance of losing \$X," where X varies from \$0 to \$11, and I proxy for loss aversion with the number of rejected lotteries. 7% of workers made inconsistent choices, accepting a lottery that is dominated by one they rejected.

I measure workers' reservation wages and their perceptions of what constitutes a "fair" wage, asking the minimum hourly wage at which they are willing to work on MTurk, and the minimum fair wage that requesters "should" pay.

The main performance measure is "Accuracy Task X", the fraction of strings entered correctly or tosses guessed correctly in stage X (in the typing task I also compute a measure of the per-character error rate, see Web Appendix B.4 for details). I try to measure how much time workers spent on their responses. There

[13]Incentivizing choices would inflate total payments, interfering with my ability to study selection effects and willingness to accept job offers for the effort task alone.

Table 2: Framing text

Experiment 1 Bonus	**Experiment 1 Penalty**
... The basic pay for the task is \$0.50. We will then randomly select one of the 50 items for checking. If you entered it correctly, the pay will be increased by \$3.00...	... 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...
Experiments 2 & 3 Bonus	**Experiments 2 & 3 Penalty**
... The pay for this task depends on your typing accuracy. We will randomly select one item for checking, and if it was entered correctly, the pay will be increased above the base pay. The base pay is \$0.50 which will be increased by \$3 if the checked item is correct...	... 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...

Note: Experiment 3 referred to "guesses" and "coin tosses" instead of accuracy and items.

are large outliers since I cannot observe how long workers were actually working on a given page of responses, only how long the page was open for, so I take the time the worker spent on the median page, multiplied by 10 to proxy the total time. Finally, to analyze beliefs, at the beginning of stage 2 workers estimated the mean accuracy rate from stage 1, a variable I label "Predicted Accuracy."

In total 1,465 workers were recruited for experiments 1 and 2, of which 693 returned for stage 2. 15 are dropped from all of the analysis, six because I have strong reasons to suspect that the same person used two MTurk accounts to participate twice[14] and nine because they scored zero accuracy in the stage 1 typing task (of the six of these who returned for stage 2, five scored zero again). Results are robust to including these nine. 398 workers were recruited for experiment 3, of which 267 completed stage 2 and 245 completed stage 3.

Randomization was stratified on the key variables on which I anticipated selection: stage 1 performance, rejected lotteries and reservation wage. In case nearby workers might know one another (for example, a couple who both work on MTurk), the treatments were randomized and standard errors clustered at the zipcode-experiment level. Web Appendix B.2 plots the CDFs of key baseline observables by framing treatment, and the associated rank-sum test p-values, confirming good balance across the distributions. Web Appendix Table B3 presents

[14] I received two pairs of near identical emails, each pair within a couple of minutes, strongly suggesting that one person was operating two accounts simultaneously. The third pair was revealed by the fact they typed identical nonsense in the second stage typing task. Such behavior is rare because all US worker accounts must be linked to a unique, verified social security number.

statistical balance tests. There is good mean balance on all characteristics with the exception of the minimum fair wage, where the difference comes from differences between experiments 1 and 2, and the number of MTurk HITs completed, where the difference is driven by outliers. The main regressions control for all baseline observables.

4 Results: experiments 1, 2 and 3

This section discusses the effect of the penalty frame on workers' willingness to accept the contract, on the types of workers who select into the contract, and on performance on the job. For much of the analysis I pool the data from experiments 1 and 2 to increase power. I then discuss the follow-up survey, the coin-toss experiment 3, and effect persistence.

4.1 Acceptance

Figure 1 graphs the rates of acceptance of the stage 2 job offer by treatment. Penalty framed contracts were much more likely to be accepted than equivalent bonus framed contracts. High fixed pay also has a large effect on the acceptance rate, high variable pay does not.

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

$$Accept_i = \beta_0 + \beta_1 * Penalty_i + \beta_2 * HighFixed_i + \beta_3 * HighVariable_i + X_i'\beta_4 + \epsilon_i$$

$Penalty$ 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). Since I do not have a group with both high fixed and variable pay, the comparison group in each case is the group with low fixed and low variable pay. X_i is a vector of controls measured in stage 1. The main specifications estimate the average effect of the penalty frame across all incentive pairs to increase power.

Table 3 presents the main results. I find that switching from bonus to penalty framing increases the acceptance rate by approximately 11 percentage points. This implies a 25% higher acceptance rate under the penalty frame than the bonus frame (the acceptance rate under the bonus frame was 42%). High fixed pay increases acceptance by around 15-16 percentage points, or around 36% (the acceptance rate in the comparison group, low fixed and variable pay, was 42%). The effect of high variable pay is positive but much smaller at 3 percentage points greater take-up, and not statistically significant.

Column (5) of Table 3 interacts the penalty treatment with the high fixed and high variable pay treatments, to estimate the differential effect of penalties under these regimes. The point estimates suggest that the effect of the penalty frame on acceptance was smaller for high fixed pay (perhaps because total take-up was higher, leaving less scope for increase) and larger for high variable pay, however neither estimate is statistically significant. In addition, the point estimate on "high variable pay" is essentially zero for workers 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.

Workers who performed better in stage 1 were significantly more likely to accept the stage 2 job offer. This is consistent with the common finding that performance pay differentially selects more able or motivated workers and is discussed further in Web Appendix B.13. Workers with a higher reservation wage were significantly less likely to accept the offer. The coefficient on "minimum fair wage" is small and 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 rejected lotteries is not predictive of acceptance, whether or not I drop workers 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 workers to be less willing to accept.

4.2 Selection

Figure 2 plots CDFs of stage 1 task performance, time spent on stage 1 task, rejected lotteries and reservation wage, comparing those who accepted the bonus frame with those who accepted the penalty frame. Interestingly, the distributions are barely distinguishable for all variables except for reservation wages, consistent with no selection on these variables. I observe suggestive evidence that the penalty contract attracted workers with higher reservation wages on average, as would be expected from the higher acceptance rate. The correlation between reservation wages and other characteristics is small.

Table 5 tests for selection effects of penalty framing by regressing the key observables on contract terms, conditional on acceptance. The coefficient on "penalty frame" is interpreted as the difference in the conditional mean of the outcome in question between penalty and bonus workers.[15] The results confirm what we saw in the graphs: the differences between bonus and penalty work-

[15]In Web Appendix Table B.3 I regress acceptance on characteristics interacted with the penalty dummy, estimating to what extent that characteristic differentially predicts acceptance under the penalty contract. The results are very similar.

ers are small and not statistically significant. Focusing on task 1 accuracy (the strongest predictor of task 2 performance), the point estimate implies 0.2 percentage points higher task 1 accuracy among penalty contract acceptors. Multiplied by the estimated coefficient on task 1 accuracy in the main performance regressions (0.72, see Table 4 column (2)), this implies less than 0.2 percentage points higher performance under the penalty contract explained by selection on task 1 performance, less than 5% of the estimated treatment effect. I discuss selection effects again when analyzing coefficient stability in the next section.[16]

4.3 Performance

The primary focus of the paper is on the effect of the penalty contract on job offer acceptance and selection, and the ideal experimental design for estimating incentive effects eliminates selection by removing the option of rejecting the contract, as in previous studies. It is nevertheless instructive to compare performance between framing treatments. First, I can check for higher performance under penalty contracts to replicate the existing literature. Second, it allows a further check for selection into penalty contracts.

Let Y_i be a measure of effort or performance. The basic regression is:

$$Y_i = \delta_0 + \delta_1 * Penalty_i + \delta_2 * HighFixed_i + \delta_3 * HighVariable_i + X_i'\delta_4 + \epsilon_i.$$

In general the estimates of δ_1, δ_2 and δ_3 will be biased by selection: if the workers that accept one 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 differential selection on observables between framing treatments, which would bias the estimate of the key coefficient of interest, δ_1. Moreover, since I have stage 1 measures of type, I can control for selection on observables by including these.

Figure 3 plots the mean performance on the stage 2 task by treatment group. I find that at each incentive level, performance is higher under the penalty than under the bonus frame (although not always statistically significantly so), consistent with the existing experimental studies. Web Appendix Figure B2 plots

[16] As for the other covariates of note, penalty acceptors were 6 percentage points more likely to be male than bonus acceptors (p=0.11), and 5 percentage points more likely to "mainly work on MTurk to earn money" (p<0.01, 93% of workers gave this response). They had completed around 10-30% more HITS in the past (sometimes significant, depending on how outliers are dealt with, note that I have imperfect balance on this variable), but had only 0.04 months more experience on MTurk (p=0.97) and 0.2 years less education (p=0.19). Finally, they were 0.2 percentage points less likely to have made inconsistent lottery choices in stage 1 (p=0.91), a plausible proxy for inattention. None of these is consequential for performance.

CDFs of the accuracy measure, my measure of the per-character error rate, and time spent, all three distributions reflect an increase in effort. Web Appendix Table B5 gives corresponding regression results.

Table 4 presents the main regressions. Accuracy under the penalty frame was 3.6 percentage points (around 0.18 standard deviations or 6% of the mean accuracy of 0.59) higher than under the bonus frame, significant at 5% without and 1% with controls. The coefficient estimate is robust to dropping workers who made inconsistent lottery choices, workers from zipcodes with multiple respondents, and outliers on the reservation and fair wage questions. Crucially, the point estimate does not change with the inclusion or exclusion of controls, consistent with the contract frame not inducing outcome-relevant selection on observables. For selection to explain the results, there would have to be a major unobserved driver of performance that is differentially selected by the penalty frame.[17]

High fixed pay increased accuracy by around 2-4 percentage points, significant at 5% when including controls. The point estimate doubles when controls are included, indicating adverse selection induced by the higher fixed pay. High variable pay increases accuracy by around 1.4-2.5 percentage points, although this is never significant at conventional levels. Column (5) interacts the penalty dummy with high fixed and high variable pay to estimate the differential effect of penalties under each financial incentive. High fixed seems to have the same effect under both frames, while high variable pay has a smaller effect under the penalty frame. Neither estimate is significant.

First stage performance strongly predicts accuracy in the second stage. A higher reservation wage is associated with poorer accuracy, while fair wage has no effect. The number of rejected lotteries is negatively associated with accuracy and significant. A one standard deviation increase in the number of rejected lotteries is associated with around 1 percentage point lower accuracy.

Table 6 finds little evidence of strong heterogeneous effects. The main effect of rejected lotteries is negative (which is possible in the extended model in Appendix A.1). However the interaction with the penalty treatment is negative (not significant), contrary to Model Prediction 2. This seems to be driven by a small number of penalty contract recipients who rejected most of the lotteries

[17]Oster (forthcoming) points out that this stability heuristic is not valid unless paired with information on R-squared movements, and provides a formula that bounds the estimate of the treatment effect using the estimates with and without controls plus assumptions on the severity of selection bias. Even under extreme assumptions ($R_{max} = 1$, $\delta = 4$), i.e. including unobservables would explain all the variation in performance and the unobservables are four times as "important" as the observables (This is highly conservative, Oster suggests setting δ equal to 1), the coefficient on Penalty barely changes, to 3.8 percentage points.

and performed poorly, see Web Appendix Figure B3.

4.4 Follow-up survey

All workers from experiment 2 were invited to complete a short survey for a fixed payment of $2. 83% did (128 of 153 bonus workers and 124 of 151 penalty workers).[18] Questions were unincentivized and conducted after the completion and payment of stage 2.

Workers were first reminded of the job offer they received in stage 2, then asked a series of questions about it. Results are presented in Table 7. Workers were asked to indicate agreement on a 1-7 scale to whether their job offer or task was fun, easy, well paid, fair, was a good motivator, earning $3.50 was achievable,[19] understandable, and whether the principal could be trusted. Results are presented in Panel A. They were then asked to what extent they agreed that the offer was attractive because of good pay, because they would be elated to receive $3.50, and because it encouraged effort, and to what extent it was unattractive because it was risky, because they would be disappointed to receive $0.50, and because it was difficult. Third, they were asked to guess the acceptance rates of workers who received the same job offer as they did, and the fraction who received the maximum pay of $3.50.

For most questions I find no significant differences between frames. However the penalty offer was rated significantly higher for good pay and more attractive due to good pay. Estimated acceptance rates and success rates were not significantly different between bonus and penalty frames, in fact penalty contract recipients thought workers were 1.4 percentage points *less* likely to receive the bonus (see Web Appendix Figure B5 for the distributions of responses to this question). Penalty workers also responded more negatively on the achievability of earning the bonus.

Workers were also shown the alternative framing of their contract and asked to rate it on various scales. I find no significant differences in ratings between bonus and penalty recipients, i.e. when asked to consider the alternative contract, penalty recipients do not on average rate the bonus contract more or less favorably than bonus recipients rated the penalty contract.[20] Experiment 6, described below, asks workers to make a direct choice between contracts.

[18] Workers who accepted in stage 2 were more likely to complete the survey (96% vs 73%, p-value < 0.001), probably reflecting that some non-participation in stage 2 will be driven by workers who did not see either of my emails.

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

[20] Additional results are presented in Web Appendix Table B6.

4.5 Experiment 3 and effect persistence

Experiment 3 was designed to shut down a number of possible mechanisms for the earlier results. It followed the same basic structure as experiments 1 and 2, but the task was changed to guessing 50 coin tosses rather than typing 50 strings, with the bonus contingent on one randomly selected toss.[21] This has two important effects. First, the task makes it much harder to hold wrong beliefs about the success probability, which is exactly 0.5. Second, it eliminates the link between effort and performance, which enables me to rule out a "taste for commitment" explanation for the findings. I also added a third stage, inviting workers back for a "Final Task," to test for experience effects.

In stage 2 the penalty contract was once again significantly more likely to be accepted than the bonus contract. 62% of bonus workers and 72% of penalty workers completed the task (difference p=0.043).[22] These results are presented in Table 8, Panel A and Web Appendix Figure B6, Panel A. The effect size, 10 percentage points, is very close to the average effect in experiments 1 and 2, though difficult to compare due to the differences in design.

Importantly, stage 1 accuracy no longer significantly predicts acceptance[23] and penalty workers did not spend significantly more time on the task[24] consistent with them understanding that effort or skill cannot affect performance on this task. I find no evidence of selection on observables into the penalty contract, see Web Appendix Table B8.

Stage 3 tested whether the popularity of the penalty persisted, by re-offering workers the same contract (now described as a "final task") one week after stage 2. Suppose workers mistakenly over-accepted the penalty contract (or under-accepted the bonus contract). Then acceptance rates in stage 3, conditional on acceptance in stage 2, should be reversed.[25] The findings are presented in Table 8,

[21] Unlike the prior task workers were told they had to submit all 50 guesses to be paid. Completion rates in experiments 1 and 2 were 93-95% and not different between frames.

[22] If I code partial completers as acceptors the figures are 67 and 75% respectively, difference p=0.073.

[23] In experiments 1 and 2 a 1 percentage point improvement in stage 1 performance is associated with a highly significant 0.3 percentage point higher acceptance rate. In experiment 3 it is associated with a nonsignificant 0.1 percentage point lower acceptance rate, though note the confidence interval is quite large. Standardizing, a 1 s.d. improvement in stage 1 performance is associated with 5.4 percentage points higher acceptance in experiments 1 and 2, and less than 1 percentage point lower acceptance in experiment 3.

[24] Mann-Whitney U p-values 0.35 and 0.80 for stages 2 and 3 respectively.

[25] For example, suppose that proportion p of workers would prefer the contract to their outside option in the absence of a framing effect, but that proportion $p_h > p$ actually accept in stage 2 under the penalty contract, and $p_l < p$ under the bonus contract. Of these, $p_h - p$ penalty workers, and no bonus workers learn that they made a mistake, and drop out in stage 3. In addition a worker randomly attrits in stage 3 with independent probability p_a. Then, conditional on accepting in stage 2, the stage 3 acceptance rate among penalty workers is

Panel B and Web Appendix Figure B6, Panel B. Yet again, the penalty contract was significantly more popular overall. Most importantly, it was more popular among those who accepted the offer in stage 2. It was also more popular among those who did not accept, and among those who accepted in stage 2, it was more popular both among those who were lucky (guessed correctly and received the bonus) and among those who were unlucky. Note however that while none of these estimated subgroup effects were negative, none are statistically significant.[26]

5 Mechanisms

5.1 Misunderstanding

There are two important ways in which workers might have misunderstood the contracts, I address each in turn. Perhaps they misunderstood and perceived the base pay as a true "base," a minimum amount below which they could not go, or perhaps they ignored the part of the contract describing the contingent pay, leading them to believe the penalty contract paid more.

Experiment 2 rephrased the job offer to address these (see Table 2). It emphasized that pay depended on performance via a bonus (or penalty), and that this would *increase pay above* or *reduce it below* the base pay respectively, i.e. it emphasized that the base pay was *not* the minimum. It also put the pay information in a single short sentence, to make it much harder to ignore part of the information. The penalty contract acceptance rate was 12 percentage points, 32%, higher ($p = 0.037$) than the bonus acceptance rate.

Additional evidence comes from experiments 3 and 4. Experiment 3 replicates the higher penalty take-up rate in stage 3 when workers have been exposed to it twice. Experiment 4 (described in detail below), elicited valuations for hypothetical contracts where the fixed pay and base pay were equal amounts, reducing scope for computational errors and misunderstanding, and finds significantly higher valuation of and implied acceptance probability for the penalty contract.

5.2 Inference

A second concern is that the bonus and penalty contracts might have induced different expectations about earnings under the contract. For example, a contract

$(1 - p_a)p/p_h$ and a higher $1 - p_a$ for bonus workers.

[26] I also examine within-task persistence of the effort effect in experiments 1 and 2, and find that the difference in performance persisted throughout the task. See Web Appendix Figure B8 and Table B15. Hossain and List (2012)'s framing effect persisted lasted several weeks.

that penalizes failure might be seen as easy (failure is unlikely) while one that rewards success is seen as hard.[27] The experiment was designed to give workers experience on the task to avoid such inferences. This section discusses additional evidence to rule out inference as an explanation for the results.

First, I attempted to measure beliefs. At the start of stage 2 of experiments 1 and 2, workers who had accepted the job offer estimated (unincentivized) the average typing accuracy rate from stage 1. They were asked about stage 1 accuracy because this estimate should only depend upon their recollection of stage 1 and the contract they received, it is not confounded by different beliefs about effort provision in stage 2 induced by the framing treatment. Bonus workers estimated a mean stage 1 accuracy of 57.4% (s.d. 19.4) and penalty workers estimated 57.9% (s.d. 18.1). The difference was not significant (p=0.75), and the full distributions of estimates are not distinguishable between bonus and penalty workers (Web Appendix Figure B4).

The strongest evidence, however, comes from Experiments 3 and 4. In experiment 3, the chance of success was transparently 50%, and yet the magnitude of the framing effect on the acceptance rate was not diminished. Experiment 4, described below, explicitly stated the probability of success at a hypothetical task (65%) and replicated the preference for penalties.

A simple back-of-the-envelope calculation illustrates how wrong beliefs would have to be to explain the results. Assuming conservatively that workers expected the typing task to take 30 minutes, the high fixed pay treatment increased hourly earnings by \$3 and the acceptance rate by around 15 percentage points, or 5 percentage points per \$1/hr. Using this figure I calculate the (risk neutral) difference in expected earnings between bonus and penalty workers required to explain the acceptance rate effects observed in Figure 1. I then compare the implied probabilities to the true distribution of success probabilities: accuracy in task 2, which had a standard deviation of 20 percentage points. The implied differences in perceived success rates between bonus and penalty workers are 3.6, 2.5, 1.3, 2.0 standard deviations for the baseline, high fixed pay, high variable pay and experiment 2 treatments. Applying the same method to the 10 minute coin toss experiment, penalty workers would need to believe their guessing accuracy was 32 percentage points higher than bonus workers did, which is 4.6 s.d. of the true, binomial accuracy distribution. Such large differences in beliefs should surely show up in the various attempts to elicit them.

[27] Bénabou and Tirole (2003) study an asymmetric information setting in which high pay signals that the task is undesirable. However, this result relies on the pay acting as a costly signal; simply altering a frame is costless.

5.3 Commitment: penalties increase effort and earnings

It is plausible that workers might like penalty contracts as a commitment device. Kaur et al. (2015) find that workers select into a strictly dominated financial incentive scheme that acts as a commitment to higher effort provision. Perhaps workers anticipate they will work harder and earn more under the penalty contract, which makes it sufficiently attractive as a commitment device that they are willing to accept the increased exposure to losses. In Web Appendix A.4 I extend the model to incorporate such a possibility.

Two pieces of evidence contradict this story. The strongest evidence is from the coin toss experiment. Since earnings are independent of effort, if workers like penalty contracts only for their motivational power, this treatment should eliminate the effect. Instead, find significantly higher take-up rate for the penalty contract in stages 2 and 3 of the experiment. This does not of course prove that commitment motives play no role, but at the very least strongly suggests that they cannot explain all of the take-up difference between bonus and penalty frames.

Second, a "meta-rational" taste for penalties does not align with actual earnings differences between treatments. For experiment 1 & 2 workers who accepted the contract I compute expected earnings as $w + eb$, where e is their realized stage 2 accuracy. I regress these measures on contract terms in Web Appendix Table B14. The penalty frame increased expected pay by 6 cents on average, and increased take-up by 11 percentage points. High fixed pay increased expected pay by \$1.53, and take-up by 16 percentage points. High variable pay increased expected pay by 91 cents but take-up by only 3 percentage points. If I compute monetary surplus, by subtracting the time spent on the task multiplied by their reservation wage, the picture is even more stark. Surplus was 29 cents *lower* on average under penalty than bonus contracts, as workers spent more time and those with higher reservation wages selected in. It was \$1.25 higher under high fixed pay than in the baseline treatment, and 83 cents higher under high variable pay. Rational expectations about earnings cannot explain such patterns in acceptance rates.

5.4 Risk seeking in the loss domain

The benchmark model assumed that reference points are choice acclimating, i.e. the reference point against which an option (contract or outside option) is evaluated is the one induced by that option. A knows that if she accepts the contract her reference point will be $w + Fb$ and if she takes the outside option it will be

$\bar{u}$, and can avoid exposure to disappointment by rejecting the job offer.

A plausible alternative is that both the contract and outside option are evaluated against reference points that are affected by the framing treatment. Intuitively, after being exposed to a high reference pay in the penalty contract, taking the outside option could feel like a loss: the framing effect "spills over" onto the outside option under consideration. This might make workers more likely to accept a penalty contract simply because the outside option is no longer as attractive. I formalize the idea in web Appendix A.5. Formally the driver is diminishing sensitivity which leads to risk-seeking (taking the risky contract) in the loss domain, akin to Tversky and Kahneman (1981)'s "Asian Disease" paradigm.

Experiment 4 tested this hypothesis by eliciting valuations of risky, gain/loss framed and safe, neutrally framed contracts side-by-side. The contracts were hypothetical typing tasks, the risky contract had \$2 fixed pay and \$2 variable pay, while the safe contract paid \$3. I elicited the maximum amount of time the worker would be willing to work under those terms. The basic idea is that if considering the framed contract shifts the reference point against which the safe contract is evaluated, that should show up in the valuation data. In fact, while the data reproduce the tendency to value penalty contracts more than bonus contracts (and to be more likely to prefer them to the outside option) I do not find any effect on the valuation of the safe contracts. See web Appendix B.8.1 for detailed results.

5.5 Salience

People sometimes overweight salient (and potentially irrelevant) information when forming judgements or evaluations. For example, in online auctions people appear to underweight shipping costs relative to base prices, so that total sale prices (winning bid plus shipping cost) are increasing in the shipping cost (Hossain and Morgan, 2006). Demand is more responsive to base prices than sales taxes, unless taxes are made salient (Chetty et al., 2009). Used-car prices are oversensitive to the left digit on the odometer (Busse et al., 2013; Lacetera et al., 2012). Reminding experienced managers and workers on a lettuce farm of their daily piece rate increased performance and shifted effort toward the incentivized dimension (Englmaier et al., forthcoming). Relatedly, choice is influenced by irrelevant anchors: valuations of a good are influenced by the distribution of possible prices (Mazar et al., 2013) and by first asking subjects whether their valuation is higher or lower than a transparently uninformative random number (Ariely et al., 2003).

A plausible explanation for the preference for penalty contracts is that workers focus their attention on the salient reference pay when evaluating whether to

accept. Thus the framing both shifts the applicable reference point when choosing effort, and also the valuation of the contract. Web Appendix A.6 suggests an extension of the model to capture this idea.

In support of this hypothesis, in the follow-up survey, penalty workers subjectively rated their contract higher for "good pay", and attractive "due to good pay", than bonus workers, despite the evidence that actual beliefs did not differ. It seems that workers formed a subjective positive assessment of the penalty contract, while objectively understanding the terms.

Although intuitive, salience as a hypothesis is difficult to formally define and therefore difficult to conclusively test.[28] I present two experiments that attempt to do so, by reducing salience in different ways. Experiment 5 simply adjusts the stage 2 invitation procedure to encourage workers to focus more on both the base and contingent pay. Experiment 6 asks workers to make a direct choice between bonus and penalty, which should make transparent the equivalence of the two. In both cases I find no difference in take-up between bonus and penalty contracts.

5.5.1 Experiment 5: Reducing Salience

Experiment 5 attempted to make the reference pay less salient to workers. It involved the same basic procedure as experiments 1 and 2: a fixed-pay typing task, a performance report, an invitation to a new task with framed incentives one week later. A slightly shorter task was used (30 rather than 50 items, eliminating the most difficult last 20 items) and the pay reduced accordingly. Payment was \$1.75 for stage 1. For stage 2 I retain the 50-50 split structure from experiment 4, with a fixed pay of \$1 and a variable pay of \$1. 797 subjects completed the first stage and were invited to the second stage. The randomization was stratified as before, balance checks are reported in web Appendix Table B9.[29]

Two different treatments were run, which made some changes to the second stage. Both sought to reduce salience to different degrees. In both treatments:

1. Workers were required to click through to the task to see the pay structure, rather than reporting the pay structure in the email. The email text is

[28] The recent literature on context-dependent choice (e.g. Bordalo et al. (2012, 2013), Kőszegi and Szeidl (2012), Cunningham (2013), Bushong et al. (2015)) formalizes the concept of salience as a property of *attributes* that depends on the distribution of those attributes in the choice set. However, it is not clear how one should apply these theories to the present context. Each theory takes for granted a canonical space of attributes. Under the most natural one (monetary outcomes) bonus and penalty contracts are identical. That is not to say that one could not come up with an attribute space that generates a penalty preference (or a bonus preference) under one or more of these theories, but I am not aware of an *a priori* compelling one.

[29] As in the earlier analysis I drop 19 from the analysis who scored zero on the first stage. This is not important for the results.

provided in web Appendix C. The goal was to make workers engage with the task before seeing the payment information.[30]

2. Workers were reminded to "please read the following carefully" at the top of the instructions.

3. Before beginning the task workers were asked to report back the payment amounts from the contract, and reminded if they entered incorrectly.[31]

The difference between the treatments was in how the payments were reported. The "conventional" treatment used the same framing as experiment 2. The "table" treatment sought to further decrease the salience of the reference pay. It:

4. reported the final payment amounts in a table (total pay if successful or unsuccessful). Bonus and penalty framing was implemented by retaining the language of "increase" or "decrease," and by reporting the reference pay first.

The exact phrasing (penalty version) was:

> 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.
>
> | | |
> |---|---|
> | Total pay if correct: | \$2 |
> | Total pay if incorrect: | \$1 |

Full regression results are documented in web Appendix B.9. Unlike the four previous experiments, neither version of the "low-salience" experiment found a higher acceptance rate for the penalty contract. In total, 64% of workers accepted the bonus contract, and 60% accepted the penalty contract; a zero difference lies well within the 90% confidence interval. The difference was slightly higher (5 percentage points versus 2.5 percentage points, not significantly different) for the conventional treatment, and average take-up was also slightly higher for the conventional treatment (65% versus 60%, not significantly different).[32]

[30] Unfortunately, due to a a survey coding error I do not observe if a worker clicked through to the task but chose not to proceed.

[31] To avoid changing the frame, in the conventional treatment they reported the base pay and the amount of the possible pay increase or decrease, while in the table treatment they reported the amount if correct or incorrect, in the same order as the instructions.

[32] In a single specification, the robustness check that drops a number of subgroups, the take-up rate of the penalty contract in the conventional treatment (but not the table treatment) is significantly lower than the bonus contract.

As before, I find no evidence of selection on observables. In addition, there was no statistically significant difference in accuracy between treatments, if anything, accuracy was lower among penalty contract recipients. Plausibly, a secondary effect of the salience manipulations was to weaken the effect of the framing on workers' reference points. Of course, the standard caveat applies–these experiments are not optimally designed to study effort choices.

Workers could submit the comprehension questions without proceeding to the task, but only 25 did so. Out of 519 submitted responses, six were incorrect.[33]

It is instructive to compare the results of experiment 5 with experiments 1 and 2, and Table 9 pools the data from these three experiments, to statistically test whether the penalty effect on take-up and performance was more positive in the high salience than the low salience experiments. Because I find little difference between the conventional and table treatments in experiment 5, I test whether experiment 5 as a whole significantly reduced the take-up or effort effect of penalties, relative to experiments 1 and 2. I find in both cases a significant difference: reducing salience significantly lowered the effect of the penalty on acceptance and performance. A note of caution though: the experiments were conducted almost three years apart, and workers were obviously not randomized into experiments. Web Appendix Table B9 tests for differences in observables between participants of experiments 1 and 2, and participants of experiment 5. There are a number of differences, most notably experiment 5 participants were more experienced. This could partly explain the change in behavior between these experiments.

Summing up, in this experiment I observe no statistically significant difference in take-up rates between bonus and penalty contracts. Notably, decreasing salience does not seem to have led to the emergence of the theoretically predicted preference for bonus contracts. But no significant effect is not the same as no effect, so it is instructive to see what effect sizes we can rule out. On average penalty contract take-up was 3.8 percentage points lower, with a 95% confidence interval of [-10.7, 3.0] (Table B10 column (1)). To contrast, the equivalent specification from pooling experiments 1 and 2 finds a 11.0 percentage point higher take-up rate under the penalty contract (95% confidence interval [5.9, 16.2]).

5.5.2 Experiment 6: Direct Choice

Experiment 6 gave workers a direct choice between an equivalent bonus and penalty contract, or the option to state "Indifferent." This plausibly reduces

[33] One from the conventional bonus treatment and five from the table bonus treatment, who gave correct values but transposed "correct" and "incorrect".

salience by making transparent the equivalence of the two contracts.

Workers were first given a short typing task (five strings, paid \$0.40) to gain experience, after which they were told their performance. Then they were asked to choose a payment scheme for typing a further ten strings. The bonus and penalty options each used the phrasing from experiment 2, with a \$0.40 fixed pay and \$0.40 variable pay paid for one randomly chosen string. They could also report indifference ("I like them the same") in which case one scheme was randomly selected for them. Indifference was always placed centrally, with the order of bonus and penalty randomized. An example screen is given in web Appendix C.8.

Eliciting incentive compatible reports of indifference is challenging, since a truly indifferent decision-maker (at least one who obeys the Independence axiom) is indifferent between choosing an option or having one randomly selected for them. To try to elicit strict preferences, half of the workers were told they would be paid an extra \$0.02 if they chose "I like them the same."

In total 295 workers completed the task, of whom 146 faced the "strict preference variant." 30% selected the bonus contract, 27% selected the penalty contract and 43% were indifferent. The difference between bonus and penalty was not significant.[34]

Strict preference elicitation seemed to make little difference: the rate of indifference was 41% in the weak and 45% in the strict preference variant, a non-significant difference ($p = 0.46$ from a two-sample test of proportions).[35]

Effort choice is particularly hard to interpret in this experiment given the prominent selection stage. For completeness, web Appendix Table B13 analyzes accuracy in the second part of the experiment. Among the 168 non-indifferent subjects, accuracy was 0.4 percentage points higher if the subject chose the penalty contract. Among the 127 indifferent subjects, effort was 4-5 percentage points higher (with/without controls) when the subject was randomly assigned a penalty contract. Neither point estimate is statistically significant.

It is interesting that many workers were not indifferent between identical bonus and penalty contracts. Evidently, multiple motives drive choice. Because the "reward" for reporting indifference was small this experiment does not tell us much about how strong these other motives are.

[34] $p = 0.54$ from a one-sample test of proportions of the null hypothesis that 50% of non-indifferent subjects prefer the penalty. Regressions controlling for stage 1 performance and order effects reach the same conclusion.

[35] The bonus contract was slightly more popular in the weak preference treatment and the penalty was slightly more popular in the strict preference treatment, neither significantly so. There is no a priori reason to expect a difference along this dimension.

5.5.3 Discussion

The evidence in this section is consistent with the hypothesis that the popularity of the penalty contract in the first four experiments can be explained by salience, and by inducing workers to focus more on the contract as a whole the take-up difference can be reduced or eliminated.

It could be that salience and framing have a tendency to go hand-in-hand. The framing effect works by making the reference outcome or state of the world salient, and inducing the decision-maker to consider other outcomes relative to the reference outcome. Then, reducing the salience of the reference outcome may also weaken the framing effect. Consistent with this view, I find a significant drop in the effort effect of penalties in experiment 5 relative to experiments 1 and 2.

One interesting implication of these results is that both manipulation of the presentation of the contracts and side-by-side comparisons were successful in eliminating the framing effect. Spiegler (2014) discusses the interplay of both effects in a general model of competitive framing.

6 Discussion and related literature

6.1 Implications and external validity

Taking the evidence as a whole, two observations are particularly notable. First, despite variation in task, presentation, worker experience and salience, none out of our six experiments has found a significant preference for bonus contracts. As argued in the introduction, this has implications for the literature that tries to understand reference-point formation, in particular suggesting that predictions made only based on loss aversion may miss other important dynamics. Strengthening external validity, these experiments took place in the type of domain and with the kinds of ranges of payments in which loss aversion has commonly been studied.

Second, relatively simple manipulations succeeded in eliminating the penalty effect. This has clear implications for external validity. Most real-world labor contracts last longer than the 20-40 minutes of these experiments, and workers are likely to focus more on their contracts, so salience might be expected to play less of a role. On the other hand, the contracts offered in this experiment were deliberately simple and transparent: a known, one-dimensional task and two payment outcomes both of which were prominently presented. Real-world contracts give more scope for salience manipulations, such as by concealing conditions in

the "small print." For example, in the tax and price salience examples part of the contract is concealed or left to the consumer to compute. Real-world contracts are also usually more complex, which increases the scope for unconscious influences on choice, as the decision-maker finds it harder to identify what it is that gives her a good feeling about the contract.[36] Finally, direct choices between contracts that are identical in every respect but framing, as in experiment 6, are rare to nonexistent.

Interestingly, I *did* find a higher take-up rate for the penalty contract in stage 3 of the coin-toss experiment. Arguably, subjects here have had more time to consider the contract as well as actual experience working under it and being paid according to their performance. Apparently salience and experience have different effects and debiasing through experience may take more than one exposure.

The results suggest a "multiple biases" view of the world. Acceptance behavior seems to be largely salience-driven, and loss aversion does not appear to have had a strong influence on contract acceptance in any experiment. Effort choices were more consistent with the predictions of loss aversion, with the caveat that the loss aversion proxy predicted behavior poorly.

Can we think of the salience effect as an example of "rational inattention?" Clearly the incentive for an MTurk worker to pay close attention to a contract for a few dollars and 20-40 minutes' work are weaker than for most workers to read their employment contracts. On the other hand, for MTurk these were not low-pay tasks, the contracts were quite transparent (and 30 minutes typing nonsense text into a browser is not an enjoyable experience!) I do not observe selection between bonus and penalty on variables expected to be correlated with inattentiveness: education, experience, or inconsistency in lottery choices (see footnote 16). My sense is that the findings are better interpreted as a cognitive mistake, one that actually makes workers worse off as argued in Section 5.3. However, to conclusively claim that this is not rational inattention I would need to observe the cost of attention, which I do not.

For what kinds of real-world contracts are these findings most relevant? A key motivation of the paper was the infrequent use of penalties in labor contracts. Given the importance of salience, we might expect to see penalties used more often in contracts that are considered infrequently, are offered in isolation or hard to compare to competitors, or are not well understood (so that it is difficult for the decision-maker to discern the influence of salience). Perhaps the best examples are from consumer choice: cellphone contracts and bank accounts that emphasize "best-case-scenario" fees, utility accounts with exit fees,

[36] See e.g. Cunningham (2014) and de Quidt and Cunningham (2016) for extensive discussion.

investment accounts with withdrawal fees, or (suggested by a referee) structured financial products that highlight a high promised return. Additionally, the fact that anticipated losses seem not to be important for choice has implications for other choices without explicit bonus/penalty features: taking on free trials failing to anticipate subsequent attachment and reluctance to cancel, or purchasing durables or housing that become difficult to part with, especially if they must be sold at a loss (Genesove and Mayer, 2001).

6.2 Related literature

Most 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. Hossain and List (2012) in a factory setting, and Church et al. (2008) and Armantier and Boly (2015) in the lab find higher effort provision under penalty framed incentives than equivalent bonus incentives. Fryer et al. (2012) test a closely related but stronger manipulation 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. Hochman et al. (2014) find similar effects with a prepayment scheme in the lab.[37] However, in a buyer-seller experiment Fehr and Gächter (2002) find that penalty-framed performance incentives reduced voluntary effort provision relative to bonuses, which they interpret as driven by fairness perceptions.

In independent laboratory work conducted simultaneously with my own, Imas et al. (forthcoming) study willingness to pay (WTP) to participate in a real-effort task to win (or avoid losing) a custom t-shirt. Consistent with my results from choice, the authors find a statistically significant 40% higher willingness to pay for the loss framed versus the gain framed task (n=85). They also find that WTP under the loss treatment is positively associated with measured loss aversion, which they argue supports the commitment mechanism since effort and earnings are increasing in loss aversion.[38] Our similar results across evaluation modes and online/lab settings complement one another.[39]

[37] de Quidt et al. (2016) do not find a significant performance effect of penalties in a study on MTurk, but their estimated effect size (0.19 s.d.) in the comparable "unannounced" treatment is of the same magnitude as I find and the sample is relatively small.

[38] They also include a treatment where subjects have no choice but to perform the task, to study performance effects in the absence of selection. They find a positive effort effect of penalties that is positively associated with the loss aversion measure, in line with Prediction 2.

[39] Brooks et al. (2014) ran an experiment exploring the limits of the effects of contract framing, and involving a participation decision. They offer three framed contracts with a low, medium and extremely high, unattainable threshold. The framing is very strong: contracts are described

This paper makes three specific contributions beyond that of Imas et al. (forthcoming). First it finds a tendency to choose the penalty contract even in the absence of the commitment motive they favor. Second, it goes further in ruling out confounding beliefs, particularly beliefs about the likelihood of success (their subjects were not given experience nor told what the performance target was). Third, it provides a manipulation, reducing salience, that is able to eliminate the tendency to choose penalties.

Another literature studies "shrouding" and complexity – efforts by firms to conceal attributes of a product or service to make it appear more attractive to consumers. Gabaix and Laibson (2006) develop the theory and show how shrouding can survive in competitive equilibrium. Brown et al. (2010) show that shrouding of shipping fees in online auctions (making them harder for consumers to learn) can be profitable. Célérier and Vallée (forthcoming) show that headline rates on consumer financial products are positively correlated with both risk and complexity, and that this increases the issuing bank's profits. The contracts in my experiments were transparent, but it is easy to create shrouded versions. For example, a penalty contract could announce just the base wage, and only disclose potential penalties when a worker applies, possibly bundled with other job features. We might expect stronger take-up effects from such contracts.

Finally, the paper relates to the literature on behavioral contract design (see Kőszegi (2014) for a review). de Meza and Webb (2007) and Herweg et al. (2010) theoretically analyze incentives for loss-averse agents without framing effects, and Just and Wu (2005) and Hilken et al. (2013) model an incentive framing problem similar to the one in this paper. Empirical papers studying loss aversion and effort provision (without framing) include Camerer et al. (1997), Goette et al. (2004), Farber (2005, 2008), Crawford and Meng (2011), Abeler et al. (2011) and Gill and Prowse (2012). The paper also fits into the smaller empirical literature on selection effects of performance pay. For example, Lazear (2000), Eriksson and Villeval (2008) and Dohmen and Falk (2011) find, as I do, that performance pay tends to select in high-ability types.

as legally binding and falling below threshold as a breach of contract. Perhaps unsurprisingly, participants were mostly unwilling to accept the extremely high threshold contract. However there is a small (not significant) increase in the acceptance rate from the low to medium threshold. Two lab experiments in the accounting literature argue for a preference for bonus contracts, but both are difficult to interpret. Luft (1994) elicits valuations for a bonus contract or a penalty contract, finding an apparent preference for bonus contracts. However, most options in the choice sets are not equivalent between between framing treatments, so choice set effects might be driving behavior. The second is Frederickson and Waller (2005), but their design introduces many additional complications that make the results difficult to interpret.

7 Conclusion

This paper presents results from six experiments studying workers' preferences over bonus and penalty contracts. Four find a higher take-up of penalty contracts which is robust to variation in task, description, evaluation mode and worker experience. Two experiments that manipulate salience are successful in eliminating the effect. In none of the six experiments do I find the theoretically predicted aversion to penalty contracts, but consistent with loss aversion, penalty framed incentives did increase performance in a typing task.

The results challenge theories that rely on workers anticipating and avoiding prospects that will expose them to painful sensations of loss, such as Herweg et al. (2010), or models where agents actively manage their reference point (e.g. Karlsson et al., 2009; Köszegi and Rabin, 2009). They point to the need for more research on the timing of formation, anticipation and determinants of reference points.

A next step would be to take these results to more natural field settings. Will loss aversion dominate when workers start to consider more richer contracts in more traditional labor markets, or will salience continue to be the key driver of choice?

Another natural question is why, in general, penalty incentives in labor contracts seem to be rarely used in practice (Baker et al., 1988; Lazear, 1991), since these results suggest that they are at worst costless for the employer, and potentially beneficial. While answering that important question is beyond the scope of this paper, candidates for future research could be that in richer environments penalties may increase multi-tasking incentives or retaliation, are damaging to group morale, or crowd out voluntary effort provision (Fehr and Gächter (2002), Grolleau et al. (forthcoming)).

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). Economic games on the internet: the effect of $1 stakes. *PloS ONE 7*(2), e31461.

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

Armantier, O. and A. Boly (2015). Framing of Incentives and Effort Provision. *International Economic Review 56*, 917–938.

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

Bell, D. E. (1985). Disappointment in decision making under uncertainty. *Operations research 33*, 1–27.

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

Benartzi, S. and R. H. Thaler (1995). Myopic Loss Aversion and the Equity Premium Puzzle. *The Quarterly Journal of Economics 110*, 73–92.

Berinsky, A. J., G. A. Huber, and G. S. Lenz (2012). Evaluating Online Labor Markets for Experimental Research: Amazon.com's Mechanical Turk. *Political Analysis 20*(3), 351–368.

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

Bordalo, P., N. Gennaioli, and A. Shleifer (2013). Salience and Consumer Choice. *Journal of Political Economy 121*(5), 803–843.

Brooks, R. R. W., A. Stremitzer, and S. Tontrup (2014). Stretch It but Don't Break It: The Hidden Risk of Contract Framing. *UCLA School of Law, Law-Econ Research Paper No. 13-22*.

Brown, J., T. Hossain, and J. Morgan (2010). Shrouded attributes and information suppression: Evidence from the field. *Quarterly Journal of Economics 125*, 859–876.

Bushong, B., M. Rabin, and J. Schwartzstein (2015). A Model of Relative Thinking. *mimeo, Harvard University*.

Busse, M. R., N. Lacetera, D. G. Pope, J. Silva-Risso, and J. R. Sydnor (2013). Estimating the Effect of Salience in Wholesale and Retail Car Markets. *American Economic Review 103*(3), 575–579.

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

Célérier, C. and B. Vallée (forthcoming). Catering to investors through security design: Headline rate and complexity. *Quarterly Journal of Economics*.

Chetty, R., A. Looney, and K. Kroft (2009). Salience and Taxation: Theory and Evidence. *American Economic Review 99*(4), 1145–1177.

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

Crawford, V. P. and J. Meng (2011). 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.

Cunningham, T. (2013). Comparisons and Choice. *mimeo, Facebook*.

Cunningham, T. (2014). Biases and Implicit Knowledge. *mimeo, Facebook*.

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

de Quidt, J. and T. Cunningham (2016). Implicit preferences inferred from choice. *CESifo Working Paper No. 5704*.

de Quidt, J., F. Fallucchi, F. Kölle, D. Nosenzo, and S. Quercia (2016). Bonus versus penalty: How robust are the effects of contract framing? *CeDEx Discussion Paper Series ISSN 1749 - 3293*.

DellaVigna, S. and D. Pope (2016). What motivates effort? evidence and expert forecasts. *NBER working paper 22193*.

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

Englmaier, F., A. Roider, and U. Sunde (forthcoming). The Role of Communication of Performance Schemes: Evidence from a Field Experiment. *Management Science*.

Ericson, K. M. M. and A. Fuster (2011). Expectations as Endowments: Evidence on Reference-Dependent Preferences from Exchange and Valuation Experiments. *The Quarterly Journal of Economics 126*, 1879–1907.

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). Is Tomorrow Another Day? The Labor Supply of New York City Cabdrivers. *Journal of Political Economy 113*, 46–82.

Farber, H. S. (2008). 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*.

Frederickson, J. R. and W. Waller (2005). Carrot or Stick? Contract Frame and Use of Decision-Influencing Information in a Principal-Agent Setting. *Journal of Accounting Research 43*, 709–733.

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., D. K. Levine, and Z. Maniadis (2012). On the robustness of anchoring effects in wtp and wta experiments. *American Economic Journal: Microeconomics 4*(2), 131–45.

Gabaix, X. and D. Laibson (2006). Shrouded Attributes, Consumer Myopia, and Information Suppression in Competitive Markets. *The Quarterly Journal of Economics 121*, 505–540.

Gächter, S., E. J. Johnson, and A. Herrmann (2010). Individual-level loss aversion in riskless and risky choices. *CeDEx discussion paper series, No. 2010-20*.

Genesove, D. and C. Mayer (2001). Loss Aversion and Seller Behavior: Evidence from the Housing Market. *Quarterly Journal of Economics 116*, 1233–1260.

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

Goette, L., D. Huffman, and E. Fehr (2004). Loss Aversion and Labor Supply. *Journal of the European Economic Association 2*, 216–228.

Grolleau, G., M. G. Kocher, and A. Sutan (forthcoming). Cheating and loss aversion: Do people cheat more to avoid a loss? *Management Science*.

Gul, F. (1991). A Theory of Disappointment Aversion. *Econometrica 59*, 667–686.

Heffetz, O. and J. A. List (2014). Is the Endowment Effect an Expectations Effect? *Journal of the European Economic Association 12*, 1396–1422.

Herweg, F., D. Müller, and P. Weinschenk (2010). 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*.

Hochman, G., S. Ayal, and D. Ariely (2014). Keeping your gains close but your money closer: The prepayment effect in riskless choices. *Journal of Economic Behavior & Organization 107*, 582–594.

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

Hossain, T. and J. A. List (2012). 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.

Imas, A., S. Sadoff, and A. Samek (forthcoming). Do People Anticipate Loss Aversion? *Management Science*.

Johnson, E. J. and D. A. Schkade (1989). Bias in utility assessments: Further evidence and explanations. *Management Science 35*(4), 406–424.

Just, D. R. and S. Wu (2005). Loss aversion and reference points in contracts. *Cornell University, Department of Applied Economics and Management working paper 127073*.

Kahneman, D., J. L. Knetsch, and R. H. Thaler (1990). Experimental Tests of the Endowment Effect and the Coase Theorem. *Journal of Political Economy 98*, 1325–1348.

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

Karlsson, N., G. Loewenstein, and D. Seppi (2009). The 'Ostrich Effect': Selective Attention to Information about Investments. *Journal of Risk and Uncertainty 38*(2), 95–115.

Kaur, S., M. Kremer, and S. Mullainathan (2010). Self-control and the development of work arrangements. *American Economic Review, Papers and Proceedings 100*(2), 624–28.

Kaur, S., M. Kremer, and S. Mullainathan (2015). Self-Control at Work. *Journal of Political Economy 123*, 1227–1277.

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

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

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

Kőszegi, B. and A. Szeidl (2012). A Model of Focusing in Economic Choice. *The Quarterly Journal of Economics 128*, 53–104.

Köszegi, B. and M. Rabin (2009). Reference-Dependent Consumption Plans. *American Economic Review 99*(3), 909–936.

Kuziemko, I., M. I. Norton, E. Saez, and S. Stantcheva (2015). How elastic are preferences for redistribution? evidence from randomized survey experiments. *American Economic Review 105*(4), 1478–1508.

Lacetera, N., D. G. Pope, and J. R. Sydnor (2012). Heuristic Thinking and Limited Attention in the Car Market. *American Economic Review 102*(5), 2206–2236.

Lazear, E. P. (1991). Labor Economics and the Psychology of Organizations. *Journal of Economic Perspectives 5*(2), 89–110.

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

List, J. A. (2004). Neoclassical Theory versus Prospect Theory : Evidence from the Marketplace. *Econometrica 72*, 615–625.

Loewenstein, G. and D. Adler (1995). A Bias in the Prediction of Tastes. *The Economic Journal 105*(431), 929–937.

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

Loomes, G. and R. Sugden (1986). Disappointment and Dynamic Consistency in Choice under Uncertainty. *The Review of Economic Studies 53*(2), 271–282.

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

Mazar, N., B. Koszegi, and D. Ariely (2013). True Context-dependent Preferences? The Causes of Market-dependent Valuations. *Journal of Behavioral Decision Making*.

Oster, E. (forthcoming). Unobservable selection and coefficient stability: Theory

and evidence. *Journal of Business & Economic Statistics*.

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

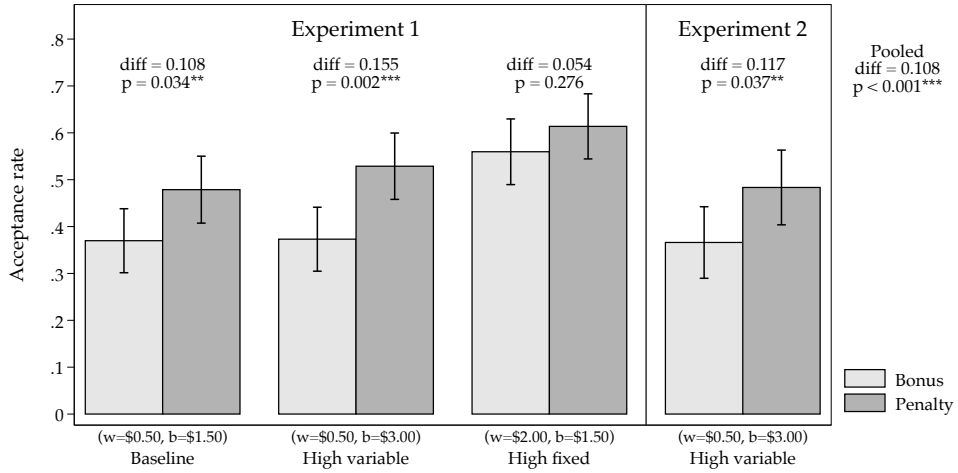
Spiegler, R. (2014). Competitive framing. *American Economic Journal: Microeconomics 6*(3), 35–58.

Tversky, A. and D. Kahneman (1981). The framing of decisions and the psychology of choice. *Science 211*(4481), 453–458.

Van Boven, L., D. Dunning, and G. Loewenstein (2000). Egocentric empathy gaps between owners and buyers: misperceptions of the endowment effect. *Journal of Personality and Social Psychology 79*(1), 66–76.

Van Boven, L., G. Loewenstein, and D. Dunning (2003). Mispredicting the endowment effect: Underestimation of owners' selling prices by buyer's agents. *Journal of Economic Behavior & Organization 51*(3), 351–365.

Figures



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

Figure 1: Acceptance Rates by treatment.

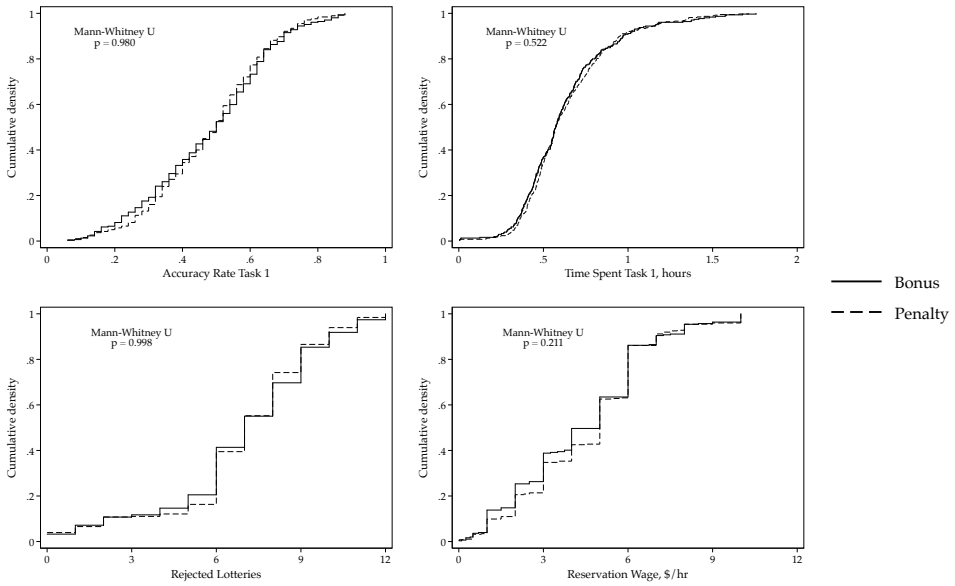
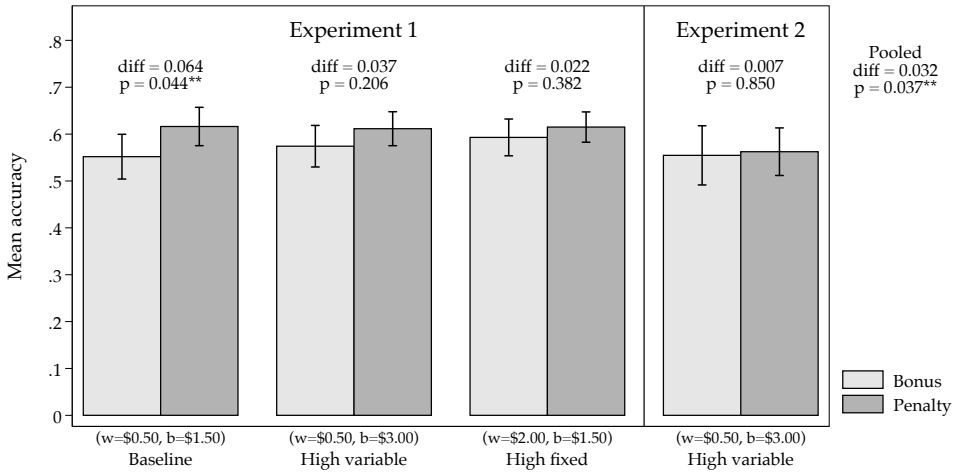


Figure 2: Comparing acceptors under Bonus and Penalty Frame. (Reservation wage and time spent trimmed at the 99th percentile.)



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

Figure 3: Accuracy by treatment.

Tables

Table 3: Acceptance decision, Experiments 1 and 2

	(1) Accepted	(2) Accepted	(3) Accepted	(4) Accepted	(5) Accepted
Penalty Frame	0.110***	0.110***	0.117***	0.104***	0.107**
	(0.026)	(0.026)	(0.029)	(0.026)	(0.050)
High Fixed Pay	0.163***	0.160***	0.151***	0.154***	0.190***
	(0.036)	(0.036)	(0.041)	(0.036)	(0.050)
Penalty * High Fixed					-0.071
					(0.071)
High Variable Pay	0.027	0.025	0.007	0.020	0.004
	(0.037)	(0.036)	(0.040)	(0.036)	(0.048)
Penalty * High Variable					0.032
					(0.062)
Accuracy Task 1		0.312***	0.141	0.337	0.343
		(0.075)	(0.352)	(0.309)	(0.309)
Accuracy Task 1 ^2			0.098	-0.055	-0.060
			(0.380)	(0.337)	(0.338)
Hours on Task 1		-0.139***	-0.198***	-0.164***	-0.165***
		(0.042)	(0.059)	(0.044)	(0.044)
Rejected Lotteries		0.005	-0.006	0.004	0.004
		(0.013)	(0.015)	(0.013)	(0.013)
Reservation wage		-0.021***	-0.013	-0.016***	-0.016***
		(0.005)	(0.009)	(0.005)	(0.005)
Fair wage		0.004	-0.006	0.005	0.006
		(0.006)	(0.009)	(0.006)	(0.006)
Experiment 2	-0.026	-0.011	-0.002	-0.012	-0.012
	(0.038)	(0.038)	(0.041)	(0.038)	(0.038)
Inconsistent lotteries				-0.065	-0.064
				(0.050)	(0.050)
Set dummies	Yes	Yes	Yes	Yes	Yes
Controls	No	No	Yes	Yes	Yes
N	1450	1448	1145	1447	1447
R-squared	0.034	0.066	0.119	0.114	0.115
Mean dep. variable	0.474	0.474	0.480	0.475	0.475

Dependent variable is a dummy indicating whether the worker accepted the job offer in stage 2. Standard errors clustered at zipcode-experiment level in parentheses. * p<0.10, ** p<0.05, *** p<0.01. Two workers are missing timing variable and one missing age variable. Column (3) drops workers 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 experiment. "Set dummies" indicate the set of strings workers typed in stage 1. "Controls" are the full set of variables collected in the stage 1 survey. "Rejected lotteries" is standardized to mean zero, standard deviation one.

Table 4: Performance on stage 2, Experiments 1 and 2

	(1) Accuracy	(2) Accuracy	(3) Accuracy	(4) Accuracy	(5) Accuracy
Penalty Frame	0.036**	0.035***	0.025*	0.036***	0.044*
	(0.015)	(0.012)	(0.014)	(0.012)	(0.027)
High Fixed Pay	0.024	0.032*	0.039**	0.038**	0.039
	(0.020)	(0.017)	(0.019)	(0.017)	(0.026)
Penalty * High Fixed					-0.001
					(0.034)
High Variable Pay	0.014	0.023	0.018	0.025	0.035
	(0.022)	(0.018)	(0.022)	(0.018)	(0.028)
Penalty * High Variable					-0.018
					(0.034)
Accuracy Task 1		0.715***	0.986***	0.996***	0.990***
		(0.037)	(0.188)	(0.172)	(0.173)
Accuracy Task 1 ^2			-0.296	-0.308*	-0.302*
			(0.190)	(0.176)	(0.178)
Hours on Task 1		-0.029	-0.043	-0.031	-0.031
		(0.023)	(0.031)	(0.024)	(0.024)
Rejected Lotteries		-0.012*	-0.017**	-0.011*	-0.011*
		(0.006)	(0.007)	(0.006)	(0.006)
Reservation wage		-0.005*	-0.008**	-0.005*	-0.005*
		(0.003)	(0.004)	(0.003)	(0.003)
Fair wage		-0.001	0.000	-0.000	-0.000
		(0.002)	(0.004)	(0.002)	(0.002)
Experiment 2	-0.038	-0.019	0.002	-0.016	-0.016
	(0.024)	(0.019)	(0.021)	(0.019)	(0.019)
Inconsistent lotteries				-0.073**	-0.073**
				(0.029)	(0.029)
Set dummies	Yes	Yes	Yes	Yes	Yes
Controls	No	No	Yes	Yes	Yes
N	687	687	550	687	687
R-squared	0.059	0.442	0.470	0.476	0.477
Mean dep. variable	0.590	0.590	0.596	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-experiment level in parentheses. * p<0.10, ** p<0.05, *** p<0.01. Column (3) drops workers 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 experiment. "Set dummies" indicate which set of strings workers typed in stages 1 and 2. "Controls" are the full set of variables collected in the stage 1 survey. "Rejected lotteries" is standardized to mean zero, standard deviation one.

Table 5: Selection, Experiments 1 and 2

	(1) Correct T1	(2) Hours on T1	(3) Rej. lotteries	(4) Res. wage	(5) Fair wage
Penalty Frame	0.002	0.006	0.004	0.245	0.111
	(0.014)	(0.022)	(0.075)	(0.194)	(0.230)
High Fixed Pay	-0.011	-0.009	-0.122	0.395	0.243
	(0.018)	(0.028)	(0.092)	(0.258)	(0.267)
High Variable Pay	-0.019	-0.026	-0.249**	0.153	0.563*
	(0.018)	(0.031)	(0.105)	(0.271)	(0.334)
Experiment 2	-0.015	0.051	0.055	0.535*	0.501
	(0.020)	(0.034)	(0.122)	(0.302)	(0.369)
N	687	687	687	687	687
R-squared	0.004	0.004	0.009	0.011	0.017
Mean dep. variable	0.481	0.635	0.032	4.587	5.896

Table regresses key observables on contract terms, conditional on contract acceptance. Standard errors clustered at zipcode-experiment level in parentheses. * p<0.10, ** p<0.05, *** p<0.01. "Rejected lotteries" standardized in the full sample to mean zero, standard deviation one.

Table 6: Stage 2 Performance: Heterogeneous effects

	Correct T1	Hours on T1	Rej. lotteries	Res. wage	Fair wage
	(1) Accuracy	(2) Accuracy	(3) Accuracy	(4) Accuracy	(5) Accuracy
Characteristic	0.745***	-0.019	-0.005	-0.004	-0.003
	(0.053)	(0.036)	(0.010)	(0.004)	(0.003)
Penalty * characteristic	-0.071	-0.024	-0.012	-0.002	0.006
	(0.075)	(0.044)	(0.013)	(0.005)	(0.004)
Set dummies	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
N	687	687	687	687	687
R-squared	0.474	0.474	0.474	0.474	0.475
Mean dep. variable	0.590	0.590	0.590	0.590	0.590

"Characteristic" corresponds to variables at top of each column. Dependent variable is accuracy in the stage 2 typing task measured as the fraction of items entered correctly. Standard errors clustered at zipcode-experiment level in parentheses. * p<0.10, ** p<0.05, *** p<0.01. "Controls" are the full set of regressors from Table 4 specification (4) excluding Accuracy Task 1 squared. "Set dummies" indicate which set of strings workers typed in stage 1. "Rejected lotteries" is standardized to mean zero, standard deviation one.

Table 7: Follow-up Survey, Experiment 2

Panel A

	The job offer or task is...							
	(1) Fun	(2) Easy	(3) Well paid	(4) Fair	(5) Motivational	(6) Trustworthy	(7) Achievable	(8) Understandable
Penalty	-0.034	-0.030	0.097**	-0.057	-0.046	-0.016	-0.069*	-0.030
	(0.042)	(0.039)	(0.040)	(0.042)	(0.038)	(0.028)	(0.036)	(0.026)
Accepted	0.154***	0.104**	0.187***	0.229***	0.205***	0.115***	0.215***	0.036
	(0.045)	(0.045)	(0.048)	(0.046)	(0.044)	(0.031)	(0.042)	(0.029)
N	252	252	252	252	252	252	252	252
R-squared	0.317	0.220	0.297	0.311	0.283	0.301	0.335	0.192
Mean Y	0.455	0.395	0.620	0.587	0.710	0.810	0.663	0.847

Panel B

	Offer attractive because...			Offer unattractive because...			Estimated aggregate behavior	
	(1) Good pay	(2) Elation of $3.50	(3) Motivational	(4) Risky	(5) Disappointment of $0.50	(6) Difficult	(7) Est. acceptance rate	(8) Est. success rate
Penalty	0.090**	0.051*	-0.041	-0.028	0.067	0.031	-0.005	-0.014
	(0.039)	(0.031)	(0.041)	(0.043)	(0.043)	(0.038)	(0.033)	(0.038)
Accepted	0.182***	0.094***	0.206***	-0.309***	-0.269***	-0.293***	0.219***	0.133***
	(0.045)	(0.036)	(0.046)	(0.049)	(0.047)	(0.045)	(0.036)	(0.044)
N	252	252	252	252	252	252	252	252
R-squared	0.310	0.251	0.306	0.373	0.346	0.329	0.347	0.211
Mean Y	0.682	0.811	0.653	0.526	0.592	0.602	0.560	0.446

Standard errors clustered at zipcode level in parentheses. * p<0.10, ** p<0.05, *** p<0.01. Dependent variables in Panels A and B are measures of agreement to statements about the job offer the worker received, measured on a 1-7 scale then transformed to a 0-1 scale. Panel C presents results from asking workers to estimate the fraction of workers 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. Regressions also control for whether the worker accepted the job offer since experience might change responses, results are very similar when this variable is excluded, or when estimated separately for acceptors and rejectors.

Table 8: Acceptance of stage 2 and 3 job offers, Experiment 3

Panel A: Stage 2

	(1) Completed	(2) Completed	(3) Completed	(4) Began	(5) Began	(6) Began
Penalty Frame	0.096**	0.100**	0.112**	0.082*	0.086*	0.096**
	(0.047)	(0.047)	(0.049)	(0.045)	(0.045)	(0.048)
Accuracy Task 1		-0.110	-0.075		-0.044	-0.064
		(0.402)	(0.401)		(0.385)	(0.385)
Rejected Lotteries		-0.000	-0.011		0.010	0.001
		(0.024)	(0.025)		(0.022)	(0.024)
Reservation wage		-0.015	0.001		-0.017	-0.003
		(0.013)	(0.013)		(0.013)	(0.013)
Fair wage		-0.003	-0.003		-0.000	-0.001
		(0.013)	(0.013)		(0.014)	(0.014)
Inconsistent lotteries			-0.055			0.083
			(0.100)			(0.093)
Controls	No	No	Yes	No	No	Yes
N	398	398	398	398	398	398
R-squared	0.010	0.017	0.104	0.008	0.016	0.106
Mean dep. variable	0.671	0.671	0.671	0.709	0.709	0.709

Panel B: Stage 3

	(1) Completed	(2) Completed	(3) Completed	(4) Completed
Penalty Frame	0.124**	0.124	0.065	0.045
	(0.049)	(0.084)	(0.054)	(0.079)
Received bonus				0.067
				(0.082)
Penalty * Received bonus				0.037
				(0.107)
N	398	131	267	267
R-squared	0.016	0.017	0.006	0.016
Mean dep. variable	0.616	0.328	0.757	0.757

Dependent variable is whether the worker completed or began the coin toss guessing task. Standard errors clustered at zipcode level in parentheses. * p<0.10, ** p<0.05, *** p<0.01. "Controls" are the full set of controls from the main specifications. "Rejected lotteries" is standardized to mean zero, standard deviation one. Time spent on task 1 was unintentionally not recorded.

Panel B column (1) includes all workers. Column (2) includes only workers who did not complete the stage 2 task. Columns (3) and (4) include only workers who completed the stage 2 task. Column (4) additionally controls for whether the worker received the bonus payment in stage 2 (i.e. guessed correctly).

Table 9: Salience: comparing experiments 1, 2 and 5

	(1) Accepted	(2) Accepted	(3) Accuracy Task 2	(4) Accuracy Task 2
Penalty Frame	0.111***	0.110***	0.034**	0.034***
	(0.026)	(0.026)	(0.015)	(0.012)
Experiment 5	0.300***	0.264***	0.017	0.045**
	(0.042)	(0.045)	(0.023)	(0.020)
Penalty * Experiment 5	-0.149***	-0.149***	-0.062***	-0.046**
	(0.043)	(0.043)	(0.024)	(0.019)
Set dummies	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes
N	2228	2223	1171	1170
R-squared	0.045	0.110	0.048	0.425
Mean dep. variable	0.526	0.526	0.586	0.586

Columns (1) and (2) dependent variable is a dummy indicating whether the worker accepted the job offer in stage 2. Columns (3) and (4) dependent variable is the fraction of items entered correctly in stage 2, conditional on acceptance. Standard errors clustered at zipcode-experiment level in parentheses. * $p<0.10$, ** $p<0.05$, *** $p<0.01$. Columns (1) and (3) control for set dummies, columns (2) and (4) include all controls from the saturated main specifications. All regressions additionally control for a dummy indicating experiment 2, and dummies for the high wage, high bonus treatments and "table" treatments.

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

FOR ONLINE PUBLICATION

Jonathan de Quidt

A Theory

Proof of Proposition 1 Define $b(F), w(F)$ implicitly by $e = e^*(b(F), F)$ and $U(e^*(b(F), F), w(F), b(F), F) = u$. I will show that $d(w(F) + eb(F))/dF > 0, \forall F \in (0, 1]$ which implies that $w(F') + eb(F') < w(F) + eb(F)$ where $F' < F$.

First, note that $w(F)+eb(F) = u+c(e)-e\mu((1-F)b(F))+\lambda(1-e)\mu(Fb(F))$ thus:

$$\begin{aligned}
&\frac{d(w(F) + eb(F))}{dF} \\
&= b[\lambda(1-e)\mu'(Fb) + e\mu'((1-F)b)] \\
&\quad + \frac{db(F)}{dF}[F\lambda(1-e)\mu'(Fb) - (1-F)e\mu'((1-F)b)] \\
&= \lambda(1-e)\mu'(Fb)\left[b + F\frac{db(F)}{dF}\right] + e\mu'((1-F)b)\left[b - (1-F)\frac{db(F)}{dF}\right]
\end{aligned}$$

Differentiating (2) we obtain:

$$\frac{db(F)}{dF} = -b\frac{\lambda\mu'(Fb) - \mu'((1-F)b)}{1 + F\lambda\mu'(Fb) + (1-F)\mu'((1-F)b)}.$$

so

$$\begin{aligned}
b + F\frac{db(F)}{dF} &= \frac{b(1 + \mu'((1-F)b))}{1 + F\lambda\mu'(Fb) + (1-F)\mu'((1-F)b)} > 0 \\
b - (1-F)\frac{db(F)}{dF} &= \frac{b(1 + \lambda\mu'(Fb))}{1 + F\lambda\mu'(Fb) + (1-F)\mu'((1-F)b)} > 0
\end{aligned}$$

and hence $\frac{d(w(F)+eb(F))}{dF} > 0$, concluding the proof.

A.1 Expectations-based reference point formation

In the basic model I assume that the reference point is entirely determined by the frame. However, Kőszegi and Rabin (2006, 2007) (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 expectations-based reference points into the model.

As discussed in the paper, I use a modification of KR's Choice acclimating Personal Equilibrium (CPE) concept.[1] In CPE, the reference point for a given choice is the rationally expected distribution of outcomes conditional on that choice. Since that does not allow for framing effects, I assume that the reference point is a weighted sum of the rational expectation and the frame. The relevant "choice" at the participation stage is between taking the outside option, or taking the contract and exerting the CPE effort level.

For simplicity I assume that utility is linear in gains and losses, that equal weight is placed on consumption and gain-loss utility (i.e. KR's η parameter equals one), and no loss aversion over effort choice. Formally, for reference point r distributed according to $H(r|e,F)$, A's utility function becomes:

$$U(G|H,e,F) = \underbrace{w + eb - c(e)}_{\text{Consumption and effort cost}} + \underbrace{e\int \mu(w+b-r)dH(r|e,F) + (1-e)\int \mu(w-r)dH(r|e,F)}_{\text{Gain-loss utility}}. \tag{4}$$

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

The marginal distribution of r is:

$$h(x|e,F) = Pr(r=x|e,F) = \begin{cases} 1-e & x = \phi w + (1-\phi)(w+Fb) \\ e & x = \phi(w+b) + (1-\phi)(w+Fb) \end{cases} \tag{5}$$
$$F \in [0,1], \phi \in [0,1).$$

As before, $w + Fb$ is the "base pay" 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.

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

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 (4) and simplifying, I can write her utility as:

$$\begin{aligned} U(e,w,b,F) = eb((2-\lambda\phi)+(\lambda-1)(1-\phi)F) + e^2(\lambda-1)\phi b \\ + w - c(e) - \lambda(1-\phi)Fb. \end{aligned} \tag{6}$$

If the appropriate second-order condition is satisfied, A's optimal effort choice is the solution to the first-order condition:

$$b((2-\lambda\phi)+(\lambda-1)(1-\phi)F) + 2e^*(b,F)(\lambda-1)\phi b - c'(e^*(b,F)) = 0. \tag{7}$$

The solution, $e^*(b,F)$ is guaranteed to be non-negative by the following assumption (which corresponds to the assumption of the same name in Herweg et al. (2010)):

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

It is straightforward to check that when A is susceptible to framing effects ($\phi < 1$) and the second-order condition $2(\lambda-1)\phi b - c''(e) < 0$ is satisfied, Predictions 1, 2 and 3 go through as in the basic model. However, the second-order condition may *not* be satisfied, for instance if b is large. The reason is that now there is positive feedback in A's effort choice: an increase from e to $e+\epsilon$ increases the reference point by $\phi\epsilon b$, which in turn induces higher effort provision. I abstract from this issue by focusing on values for b smaller than some threshold $\bar{b}$, such that $c''(e) > 2(\lambda-1)\phi\bar{b}, \forall e \in [0,1]$. If the cost of effort is quadratic, $c(e) = \gamma_1 + \gamma_2 e + \gamma_3 e^2$, the condition becomes $\gamma_3 > 2(\lambda-1)\phi\bar{b}$.

As an aside, I note that in this version of the model, it is possible for e^* to decrease in λ (in the basic model it is weakly increasing). Differentiating (7), we obtain:

$$\frac{de^*}{d\lambda} = \frac{b(\phi(2e^*-1)+(1-\phi)F)}{c''(e^*) - 2(\lambda-1)\phi b}.$$

Hence if e^* is small enough and ϕ is large enough, the expression can be negative. The intuition is similar to that just discussed. Increasing λ increases the disutility of a given loss experienced when unsuccessful. This may discourage effort provision, since higher effort increases the reference point, increasing the size of the loss. The higher is ϕ, the stronger the effect of e on the reference point.

Proposition 1 also holds in this setting.

Proof of Proposition 1 for extended KR preferences As before, define $b(F), w(F)$ implicitly by $e = e^*(b(F),F)$ and $U(e^*(b(F),F), w(F), b(F), F) = u$.

I will show that $\frac{d(w(F)+eb(F))}{dF} > 0, \forall F \in (0,1]$ which implies that $w(F')+eb(F') < w(F)+eb(F)$ for $F' < F$. Differentiating (7) we obtain:

$$\frac{db(F)}{dF} = -\frac{b(\lambda-1)(1-\phi)}{2-\lambda\phi+(\lambda-1)(1-\phi)F+2e(\lambda-1)\phi}.$$

From $U(e, w(F), b(F), F) = u$ we obtain

$$\begin{aligned} &w(F)+eb(F) \\ &= u + c(e) - eb((1-\lambda\phi)+(\lambda-1)(1-\phi)F) - e^2(\lambda-1)\phi b + \lambda(1-\phi)Fb. \end{aligned}$$

thus

$$\begin{aligned} &\frac{d(w(F)+eb(F))}{dF} \\ &= b(1-\phi)(e+\lambda(1-e)) \\ &\quad + \frac{db(F)}{dF}[(1-\phi)F(e+\lambda(1-e)) + \lambda\phi e(1-e) - e(1-\phi e)] \\ &= \frac{b(1-\phi)[(2-\lambda\phi)(e+\lambda(1-e)) + e(\lambda-1)(1+\phi e+\lambda\phi(1-e))]}{2-\lambda\phi+(\lambda-1)(1-\phi)F+2e(\lambda-1)\phi} > 0. \end{aligned}$$

Hence $\frac{d(w(F)+eb(F))}{dF} > 0$, concluding the proof.

A.2 Reference-dependent effort provision

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

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

I do not propose this mechanism as a possible explanation for the main finding that penalties are more popular than bonuses, because I feel there is no a priori reason to think that a penalty frame over payoff outcomes would increase or decrease the effort reference point, and in particular that this effect would dominate the effect of loss aversion over consumption, which is the domain specifically tar-

geted by the intervention. The contract simply paid for one randomly selected item and did not specify a target overall effort level, unlike in the piece-rate incentive schemes used in other framing experiments.

I assume that gain-loss disutility in effort enters linearly, weighted by a parameter η^e, and is equal to $\lambda\eta^e(e - e^r(F))$ when $e \geq e^r(F)$ and $\eta^e(e - e^r(F))$ when $e < e^r(F)$.[2] A's utility function is now:

$$\begin{aligned} U(e, w, b, F) = w + eb - c(e) + e\mu((1-F)b) - \lambda(1-e)\mu(Fb) \\ - \eta^e(e - e^r(F))(\lambda(1 - \mathbb{1}[e < e^r(F)]) + \mathbb{1}[e < e^r(F)]). \end{aligned} \tag{8}$$

$\mathbb{1}[x]$ is an indicator function that takes value one when $x \geq 0$ and zero otherwise. Differentiating U with respect to F yields $\frac{\partial U}{\partial F} = -eb\mu'((1-F)b) - \lambda(1-e)b\mu'(Fb) + \eta^e \frac{de^r(F)}{dF}(\lambda(1 - \mathbb{1}[e < e^r(F)]) + \mathbb{1}[e < e^r(F)])$. This expression is negative (i.e. utility is decreasing in F) provided $\frac{de^r(F)}{dF}$ is sufficiently small.

To illustrate, suppose that gain-loss utility is linear in money: $\mu(x) = \eta^c x$. Then $\frac{\partial U}{\partial F} = -\eta^c b(e + \lambda(1-e)) + \eta^e \frac{de^r(F)}{dF}(\lambda(1 - \mathbb{1}[e < e^r(F)]) + \mathbb{1}[e < e^r(F)])$. A simple sufficient condition for this expression to be strictly negative is $\frac{de^r(F)}{dF} < \frac{\eta^c b}{\eta^e \lambda}$. Note that $\frac{dr}{dF} = b$, where r is the monetary reference point $w + Fb$. Therefore the condition can be written as $\frac{de^r(F)}{dF} / \frac{dr(F)}{dF} < \frac{\eta^c}{\eta^e \lambda}$. In other words, the effect of the frame on the reference point for effort, weighted by the strength of gain-loss utility in effort should not be too strong relative to the effect on the reference point for money weighted by the strength of gain-loss utility in money.

As noted above, I lack a theory of how $e^r(F)$ should depend upon F. However it seems intuitive that a manipulation that alters the framing of the financial terms of the contract should most strongly affect the reference point for money (i.e., $\frac{de^r(F)}{dF}$ is relatively small), hence I do not find reference dependence in effort a persuasive explanation for the popularity of the penalty contract. Moreover, the penalty contract was significantly more popular than the bonus contract in the coin toss version of the experiment. Since performance did not depend on effort in that experiment (and effort as proxied by time spent was not significantly different between bonus and penalty treatments), it is hard to see how reference dependence in effort could explain the penalty contract's popularity.

For completeness, A's optimal effort choice is as follows:

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

[2] Note that since effort is a non-stochastic choice, A will be either always above, always below or exactly at her effort reference point.

where e^H solves $b+\mu((1-F)b)+\lambda\mu(Fb)-c'(e^H)-\eta^e$, and e^L solves $b+\mu((1-F)b)+\lambda\mu(Fb)-c'(e^L)-\eta^e\lambda$.

Assume that e^H and e^L are strictly increasing in F, which holds when μ is linear or the condition in footnote 8 in the main text is satisfied. Then a) if $e^r(F)$ is increasing in F, then e^* is everywhere increasing in F; b) if $e^r(F)$ does not depend on F, then e^* is increasing in F for $e^H \leq e^r(F)$ or $e^L \geq e^r(F)$ and flat when $e^L < e^r(F) < e^H$, and lastly c) if $e^r(F)$ is *decreasing* in F, then e^* is first increasing, then decreasing, then increasing again.

A.3 Preferred Personal Equilibrium

The benchmark model and extension outlined above apply an extension of KR's Choice Acclimating Personal Equilibrium (CPE) concept, which cannot explain the preference for the penalty contract. In this section I show that KR's leading alternative solution concept, Preferred Personal Equilibrium (PPE), extended to allow for framing effects, also does not allow a worker to accept the penalty contract but reject the bonus contract.

Let $U(C_i|r)$, $i \in \{B,P\}$, denote the benchmark model utility of the contract evaluated against reference point r, with B,P indicating bonus or penalty framing. $U(\bar{u}|r)=\bar{u}+\mathbb{1}[\bar{u}\geq r]\mu(\bar{u}-r)-(1-\mathbb{1}[\bar{u}\geq r])\lambda\mu(r-\bar{u})$ is the utility of the outside option, $\bar{u}$.

The worker either faces the choice set $\{C_P,\bar{u}\}$, or $\{C_B,\bar{u}\}$. The penalty contract induces reference point $r(C_P)=w+b$, the bonus contract induces reference point $r(C_B)=w$ and the outside option induces reference point $r(\bar{u})=\bar{u}$.

A choice $x\in\{x,y\}$ is a *personal equilibrium* (PE) if it is chosen when the reference point is that induced by x, i.e. $x\in\{x,y\}$ is a PE if $U(x|r(x))\geq U(y|r(x))$. It is a *preferred personal equilibrium* (PPE) if it is a PE and either y is not a PE, or y is a PE but x is the (weakly) preferred PE, i.e. $U(x|r(x))\geq U(y|r(y))$.

Of course, the framing only affects utility via the reference point, i.e. $U(C_B|r)=U(C_P|r)$. It will therefore be convenient to suppress the subscript when the relevant reference point is explicitly stated.

The goal is to show that it cannot be that both C_P is a PPE from $\{C_P,\bar{u}\}$, and that $\bar{u}$ is the unique PPE from $\{C_B,\bar{u}\}$.
C_P is a PPE from $\{C_P,\bar{u}\}$ if:

(a) it is a PE: $U(C|w+b)\geq U(\bar{u}|w+b)$ and

(b1) $\bar{u}$ is not a PE: $U(\bar{u}|\bar{u})<U(C|\bar{u})$, or

(b2) $\bar{u}$ is a PE, but C_P is the (weakly) preferred PE: $U(\bar{u}|\bar{u}) \geq U(C|\bar{u})$ and $U(\bar{u}|\bar{u}) \leq U(C|w+b)$.

$\bar{u}$ is the unique PPE from $\{C_B, \bar{u}\}$ if:

(c) $\bar{u}$ is a PE: $U(C|\bar{u}) \leq U(\bar{u}|\bar{u})$ and

(d1) accepting the bonus contract is not a PE: $U(C|w) < U(\bar{u}|w)$ or

(d2) accepting the bonus contract is a PE, but the outside option PE yields strictly higher utility: $U(\bar{u}|w) \leq U(C|w)$ and $U(\bar{u}|\bar{u}) > U(C|w)$.

First, we note that $\bar{u} < w + b$ otherwise C_P cannot be a PE from $\{C_P, \bar{u}\}$, in which case (a) cannot hold. Second, $\bar{u} > w$ otherwise $\bar{u}$ cannot be a PE from $\{C_B, \bar{u}\}$, in which case (c) cannot hold.

$w + b > \bar{u} > w$ implies that $U(x|w) > U(x|\bar{u}) > U(x|w+b)$ for any prospect x. This means that (b2) cannot hold, since it requires $U(C|\bar{u}) \leq U(\bar{u}|\bar{u}) \leq U(C|w+b)$, but $U(C|\bar{u}) > U(C|w+b)$. Similarly, (d2) cannot hold, since it requires $U(\bar{u}|\bar{u}) > U(C|w) \geq U(\bar{u}|w)$, but $U(\bar{u}|w) > U(\bar{u}|\bar{u})$.

Thus we need to check if (a), (b1), (c), and (d1) can be satisfied. However, (c) contradicts (b1). Hence, just as CPE is inconsistent with the relative popularity of the penalty contract, so is PPE.

A.4 Penalties as a commitment device

The model in this section relates to the goal-setting literature, e.g. Koch et al. (2014) and Golman and Loewenstein (2012), and to multiple-selves models such as Thaler and Shefrin (1981).

Suppose that if she accepts the contract, A's "Doer" self chooses effort by maximizing (1). However, when deciding whether to accept the contract, A's "Planner" self evaluates it according to the following modified utility function:

$$V(e^*(b,F), w, b, F) = w + e^*(b,F)b - \beta c(e^*(b,F)) + e^*(b,F)\mu(b). \qquad (9)$$

Where $e^* = e^*(b,F)$ from (2) and $\beta \leq 1$. (9) is maximized at $e^* = e^{FB}(b)$ which solves the first order condition $b + \mu(b) - \beta c'(e^{FB}(b)) = 0$ and does not depend on the frame.

(9) differs from (1) in three ways. First, it assumes that A's reference point when evaluating the contract is w, i.e. she does not yet feel endowed with Fb. Second, she anticipates that Doer will update her reference point and exert effort $e^*(b,F)$. Third, she weights Doer's cost of effort by $\beta \leq 1$, reflecting her self-control problem. All else equal, Planner would like to exert more effort than

Doer. Thus, if $\beta < 1$ she expects to exert suboptimal effort under the pure bonus contract ($e^{FB}(b) > e^*(b,0)$). Planner prefers the pure penalty contract to the pure bonus contract if $V(e^*(b,1),w,b,1) > V(e^*(b,0),w,b,0)$. A simple sufficient condition is $e^{FB}(b) \geq e^*(w,b,1)$ or $\frac{b+\mu(b)}{b+\lambda\mu(b)} \geq \beta$. Increasing β or λ tightens this condition by increasing the likelihood of over-provision of effort under the penalty contract.

A.5 Risk seeking in the loss domain

For simplicity, consider the coin toss experiment where $e = 0.5$, and normalize the cost of effort to zero. Under the contract, the reference point is $w + Fb$, while under the outside option it is $r_u = (1-\psi)\bar{u} + \psi(w+Fb)$. $\psi \in [0,1]$ captures the extent to which the outside option is influenced by the treatment. Total utility under the outside option is now $\bar{U}(w,b,F) = \bar{u} + \mathbb{1}[\bar{u} \geq r_u]\mu(\bar{u} - r_u) - (1 - \mathbb{1}[\bar{u} \geq r_u])\lambda\mu(r_u - \bar{u})$. Assume that $\bar{u}$ is non-stochastic and lies between w and $w+b$ (since the mechanism relies on risk-seeking in the loss domain, it is harder to generate a preference for penalties when the outside option is also risky).

The benchmark model with choice acclimating reference points corresponds to $\psi = 0$. $\psi = 1$ captures the classic "Asian Disease" paradigm (Tversky and Kahneman, 1981), where participants are offered a choice between two alternatives, both of which are framed against the same reference point.[3] My experiments only manipulated the framing of the contract, not the outside option, so it seems reasonable to assume $\psi < 1$.

I want to see if it is possible that $U(w,b,1) > \bar{U}(w,b,1)$ and $U(w,b,0) < \bar{U}(w,b,0)$. For $\psi = 1$, this is equivalent to checking whether accepting the penalty contract is a Personal Equilibrium, and accepting the bonus contract is not, see Web Appendix A.3 for further discussion. The first expression is equal to $w + \frac{b}{2} - \frac{\lambda}{2}\mu(b) > \bar{u} - \lambda\mu(\psi(w+b-\bar{u}))$, and the second is $w + \frac{b}{2} + \frac{1}{2}\mu(b) < \bar{u} + \mu(\psi(\bar{u}-w))$. Rearranging, we need to check if it is possible that $\lambda\left[\frac{1}{2}\mu(b) - \mu(\psi(w+b-\bar{u}))\right] < w + \frac{b}{2} - \bar{u} < \mu(\psi(\bar{u}-w)) - \frac{1}{2}\mu(b)$.

Proposition 2. *Provided $\mu'' < 0$ and ψ is sufficiently large there exists an interval of values for $\bar{u}$, containing $\bar{u} = w + \frac{b}{2}$, such that $\bar{U}(w,b,0) > U(w,b,0) > U(w,b,1) > \bar{U}(w,b,1)$.*[4]

[3] The classic example considers an "Asian disease" that might kill 600 people, and offers the choice between 200 lives saved (400 deaths) and a 1/3 probability of 600 lives saved (2/3 probability of 600 deaths). The typical finding is risk aversion in the gain domain and risk seeking in the loss domain.

[4] Proof: at $\psi = 1$, $\bar{u} = w + \frac{b}{2}$ the condition reduces to $\lambda\left[\frac{1}{2}\mu(b) - \mu(\frac{b}{2})\right] < 0 < \mu(\frac{b}{2}) - \frac{1}{2}\mu(b)$ which is strictly satisfied if and only if $\mu'' < 0$. Then the ranges for ψ and $\bar{u}$ follow by an open set argument.

In other words, it is possible to generate a revealed preference for the penalty contract. The result is actually driven by diminishing sensitivity, which renders the utility function convex in the loss domain, leading to risk seeking in the loss domain. Since the outside option is riskless, it is relatively attractive in the gain domain and unattractive in the loss domain.

How plausible is this explanation for the result? The closer to linear is μ and the smaller is ψ, the smaller will be the range of values for $\bar{u}$ that generate the observed revealed preference pattern. The results of experiment 4 suggest that ψ is small.

A.6 Salience model

Salience can be captured by the following simple extension of the model. Instead of expected pay plus gain-loss utility, A evaluates the contract as the reference point plus gain-loss utility. Focusing on pure bonus and penalty contracts for simplicity, under the bonus contract A's utility is:

$$U(e, w, b, 0) = w + e\mu(b) - c(e)$$

under the penalty contract it is:

$$U(e, w, b, 1) = w + b - \lambda(1 - e)\mu(b) - c(e).$$

She prefers the penalty contract if $b - \lambda(1 - e)\mu(b) > e\mu(b)$, or $b - \lambda\mu(b) + e\mu(b)(\lambda - 1) > 0$. If $\lambda \geq 1$, a sufficient condition is $b > \lambda\mu(b)$, i.e. diminishing sensitivity more than offsets loss aversion. She exerts higher effort under the penalty contract if $\lambda > 1$, i.e. if she is loss averse, the same condition as in the original model.

B Additional Results

B.1 Summary statistics

Table B1: Performance on effort tasks, Experiments 1, 2 and 5

	N	Median	Mean	s.d.
Stage 1				
Accuracy Task 1	2228	0.44	0.44	0.18
Errors per item Task 1	2228	0.032	0.047	0.074
Hours on Task 1	2225	0.48	0.52	0.32
Stage 1, Rejectors only				
Accuracy Task 1	1057	0.42	0.42	0.18
Errors per item Task 1	1057	0.036	0.054	0.090
Hours on Task 1	1054	0.53	0.57	0.34
Stage 1, Acceptors only				
Accuracy Task 1	1171	0.47	0.46	0.17
Errors per item Task 1	1171	0.031	0.041	0.054
Hours on Task 1	1171	0.43	0.48	0.30
Stage 2				
Predicted Mean Round 1 Accuracy	686	0.60	0.58	0.19
Accuracy Task 2	1171	0.62	0.59	0.20
Errors per item Task 2	1171	0.018	0.064	0.17
Hours on Task 2	1170	0.45	0.56	0.99

"Accuracy" is the fraction of items a worker entered correctly. "Errors per item" is 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. "Predicted Mean stage 1 Accuracy" is the worker'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?" This measure was not collected in experiment 5. Timing data is missing for three observations in stage 1 and one in stage 2, likely due to JavaScript errors. One worker did not report a predicted stage 1 accuracy.

Table B2: Summary statistics from stage 1 survey, Experiments 1, 2, 3 and 5

	N	Mean	s.d.
Loss Aversion			
Rejected Lotteries	2626	6.85	2.85
Inconsistent Lottery Choices	2626	0.07	0.25
Reservation Wage			
Reservation wage, $/hr	2626	5.14	2.75
Minimum fair wage, $/hr	2626	6.41	2.74
MTurk Experience			
Hours working on MTurk per week	2626	16.5	14.5
Typical MTurk earnings, $100/week	2626	1.03	19.7
Earnings/hours, $/hr	2624	5.89	54.9
MTurk HITs completed	2626	9577	46369
Months of experience on MTurk	2626	13.2	27.4
Mainly participate in research HITs	2626	0.79	0.41
Work on MTurk mainly to earn money	2626	0.92	0.27
Demographics			
Age in 2013	2624	32.6	10.8
Male	2626	0.48	0.50
Household Income	2626	40223	28891
Zipcode cluster	2626	0.11	0.31
Employment Status			
Full time	2626	0.41	0.49
Part time	2626	0.13	0.33
Self employed	2626	0.11	0.31
Full time MTurk worker	2626	0.11	0.32
Unemployed	2626	0.10	0.31
Student	2626	0.10	0.30
Other	2626	0.04	0.21
Education			
Less than High School	2626	0.00	0.07
High School / GED	2626	0.11	0.31
Some College	2626	0.31	0.46
2-year College Degree	2626	0.12	0.33
4-year College Degree	2626	0.36	0.48
Masters Degree	2626	0.08	0.27
Doctoral/Professional Degree	2626	0.02	0.14

"Rejected lotteries" is the number of 50-50 win-lose lotteries that workers report they would be unwilling to play. "Mainly participate in research HITs" indicates workers who report mostly working on HITs posted by researchers. "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 worker in the same experiment reported the same zipcode.

B.2 Balance checks

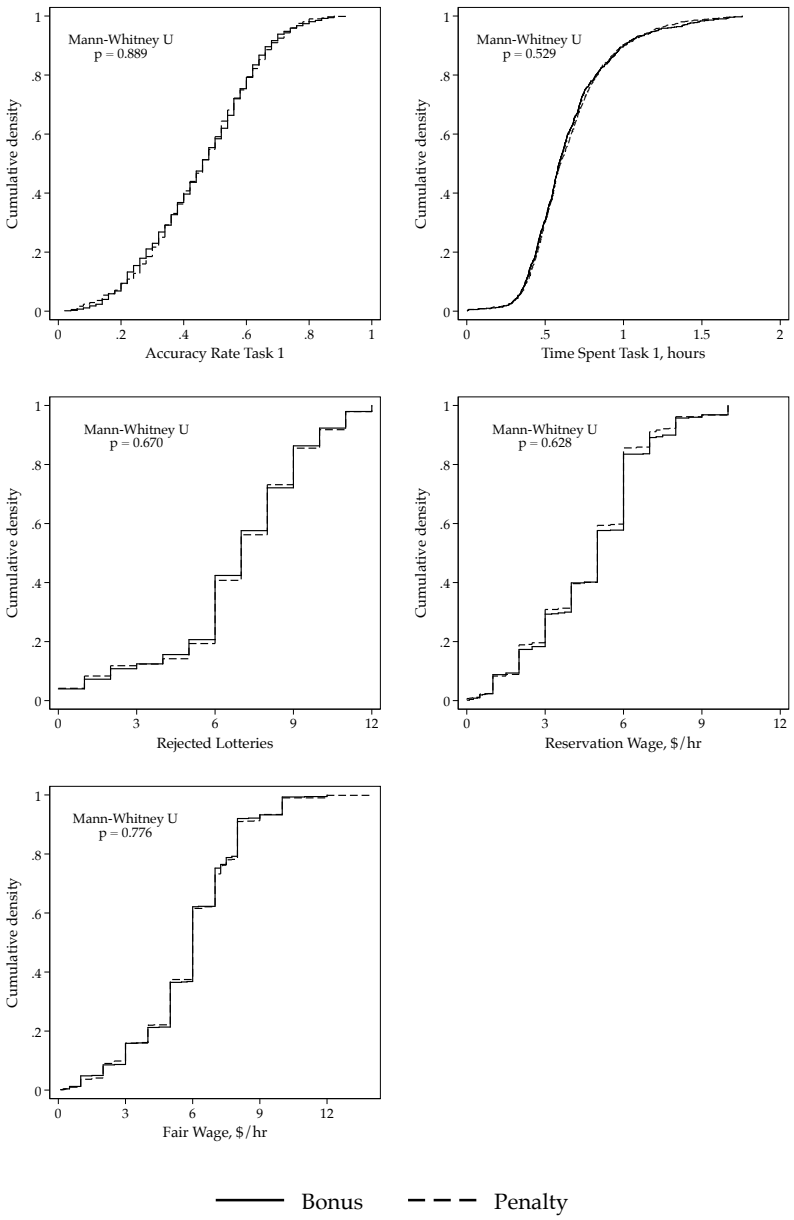


Figure B1: Balance between bonus and penalty treatments, experiments 1 and 2. Reservation wage trimmed at the 99th percentile.

Table B3: Balance Check

	Experiments 1 and 2										Experiment 3	
	Joint		Groups 0 & 1		Groups 2 & 3		Groups 4 & 5		Groups 6 & 7		Groups 8 & 9	
	F-stat	p	Diff	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	-0.00	0.84
Hours on Task 1	0.51	0.83	0.01	0.76	-0.00	1.00	0.01	0.68	-0.04	0.28	-[a]	-[a]
Rejected Lotteries	0.53	0.81	0.31	0.26	-0.08	0.79	-0.23	0.39	-0.17	0.60	-0.04	0.89
Inconsistent Lottery Choices	0.62	0.74	0.01	0.60	0.02	0.42	-0.01	0.80	0.01	0.81	-0.00	0.94
Reservation wage, $/hr	0.65	0.72	0.07	0.77	-0.02	0.93	0.10	0.77	0.06	0.84	-0.27	0.22
Minimum fair wage, $/hr	2.87	0.01***	0.08	0.73	0.05	0.88	-0.38	0.14	0.16	0.59	-0.01	0.98
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	-2.19	0.12
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	-0.13	0.08*
MTurk HITs completed	1.85	0.07*	-7499.66	0.09*	3609.55	0.07*	460.60	0.71	-1966.82	0.50	-5237.00	0.06*
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	0.38	0.80
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	-0.00	0.92
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	0.02	0.53
Age in 2013	0.73	0.65	1.38	0.18	-0.80	0.45	-0.90	0.42	0.53	0.69	-1.01	0.38
Male	0.93	0.48	0.08	0.12	-0.00	0.93	0.01	0.92	-0.01	0.81	-0.05	0.29

[a] Timing data were unintentionally not collected in the coin toss experiment task 1.

"Joint" reports the F-statistic and p-value from a joint test of the significance of the set of group dummies in explaining each relevant baseline variable, pooling data from experiments 1 and 2. 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, and groups correspond to those in Table 1. Significant difference in joint test of "minimum fair wage" driven by differences between experiments 1 and 2. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

B.3 Selection: alternative specification

Table B4: Selection: alternative specification

	(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 worker accepted the job offer in stage 2. Estimates from OLS linear probability model. Standard errors clustered at zipcode-experiment level in parentheses. * p<0.10, ** p<0.05, *** p<0.01. "Controls" are the full set of regressors from Table 3 specification (4) excluding Accuracy Task 1 squared. "Set dummies" indicate which set of strings workers typed in stage 1. Interaction variables are demeaned to stabilize interaction coefficients. "Rejected lotteries" is standardized to mean zero, standard deviation one.

B.4 Alternative performance and effort measures

In addition to per-string accuracy, I compute a measure of per-character accuracy, "Scaled Distance Task X." For each text string I compute the Levenshtein distance (the minimum number of single character insertions, deletions, or swaps needed to convert string A into string B) between the worker's response and the correct answer, and divide by the length of the correct answer, then average over all answers. This then roughly corresponds to the probability of error per character. Blank responses are coded as 1 (note that this does mean a worker who entered, say "abc" when the correct response was in fact "de" would score 1.5, i.e. worse

than had they left the answer blank). The regressions and figure use the natural log of this measure since it is is heavily skewed.

Table B5: Performance/effort on stage 2, alternative measures

	(1) Log distance	(2) Log distance	(3) Time spent	(4) Time spent
Penalty Frame	-0.205**	-0.214***	0.052*	0.026
	(0.090)	(0.080)	(0.027)	(0.023)
High Fixed Pay	-0.251**	-0.320***	0.015	0.025
	(0.119)	(0.113)	(0.036)	(0.031)
High Variable Pay	-0.048	-0.064	0.028	0.037
	(0.133)	(0.130)	(0.040)	(0.033)
Accuracy Task 1				0.173
				(0.383)
Accuracy Task 1 ^2				-0.105
				(0.382)
Log Errors per item Task 1		0.853***		
		(0.067)		
Hours on Task 1		0.277		0.736***
		(0.168)		(0.067)
Experiment 2	0.160	0.023	0.005	-0.036
	(0.149)	(0.128)	(0.046)	(0.036)
Set dummies	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes
N	687	687	680	680
R-squared	0.075	0.342	0.027	0.399
Mean dep. variable	-3.825	-3.825	0.689	0.689

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 workers above the 99th percentile for time spent. Standard errors clustered at zipcode-experiment level in parentheses. * p<0.10, ** p<0.05, *** p<0.01. "Set dummies" indicate which set of strings workers typed in stages 1 and 2. "Controls" are the full set of controls from the main specifications.

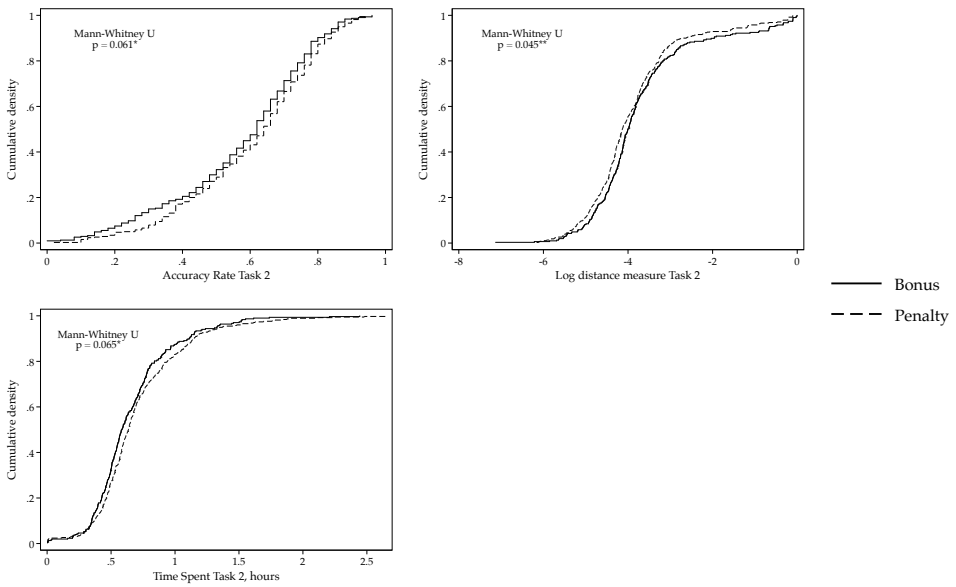
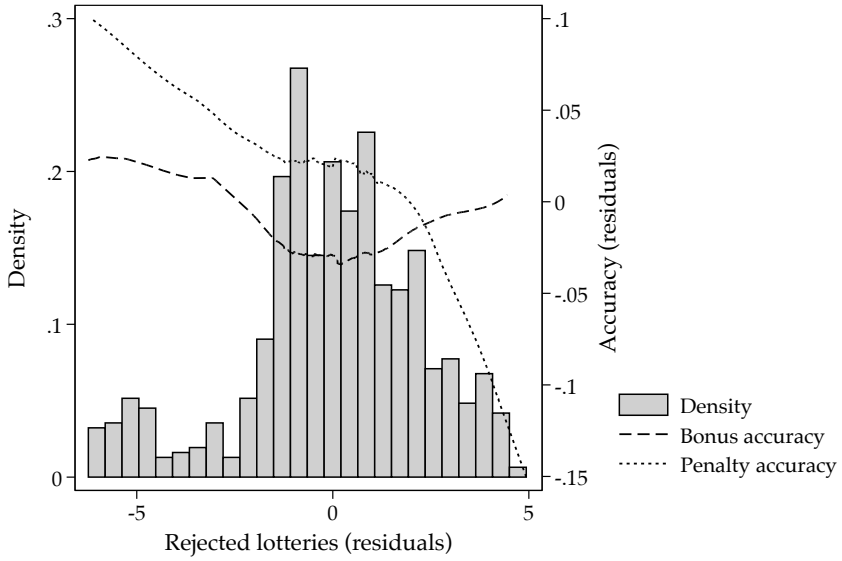


Figure B2: Performance and effort measures, comparing bonus vs penalty frame. Time spent is trimmed at the 99th percentile.

B.5 Relationship between rejected lotteries and performance



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

Figure B3: Relationship between Rejected Lotteries and performance.

B.6 Follow-up survey, other results

Table B6: Follow-up survey auxiliary results, Experiment 2

Panel A

	(1) Accept again	(2) Accept again	(3) WTA	(4) WTA
Penalty	-0.013	0.019	-0.839	-0.059
	(0.039)	(0.043)	(0.906)	(0.302)
Accepted	0.390***		-1.453	
	(0.043)		(1.162)	
Received $3.50		0.222***		-0.667
		(0.085)		(0.463)
N	252	124	252	124
R-squared	0.426	0.539	0.143	0.446
Mean Y	0.640	0.837	3.457	2.885

Panel B

	Compared to my contract the other contract is more...					
	(1) Attractive	(2) Fair	(3) Motivating	(4) Generous	(5) Trustworthy	(6) Achievable
Penalty	0.000	-0.013	0.033	-0.043	0.002	-0.022
	(0.039)	(0.035)	(0.034)	(0.035)	(0.032)	(0.035)
Accepted	-0.076*	-0.027	-0.017	-0.027	-0.044	-0.062
	(0.043)	(0.039)	(0.036)	(0.040)	(0.035)	(0.041)
N	252	252	252	252	252	252
R-squared	0.200	0.192	0.191	0.228	0.190	0.170
Mean Y	0.462	0.463	0.581	0.469	0.472	0.451

Standard errors clustered at zipcode level in parentheses. * p<0.10, ** p<0.05, *** p<0.01. Dependent variable in Panel A columns (1) and (2) is reported willingness to accept the same job again, on a 1-5 scale (definitely would not to definitely would), again standardized. Dependent variable in Panel A columns (3) and (4) is the fixed amount that would make workers indifferent between performing the task under their contract or in return for the fixed amount. This measure was very noisy. Dependent variables in Panel B are measures of agreement to statements about the job offer the worker *did not receive*, measured on a 1-5 scale then transformed to a 0-1 scale. All regressions include the full set of controls from the main specifications. Regressions also control for whether the worker accepted the job offer since experience might change responses, results are very similar when this variable is excluded, or when estimated separately for acceptors and rejectors.

B.7 Workers' predictions of performance

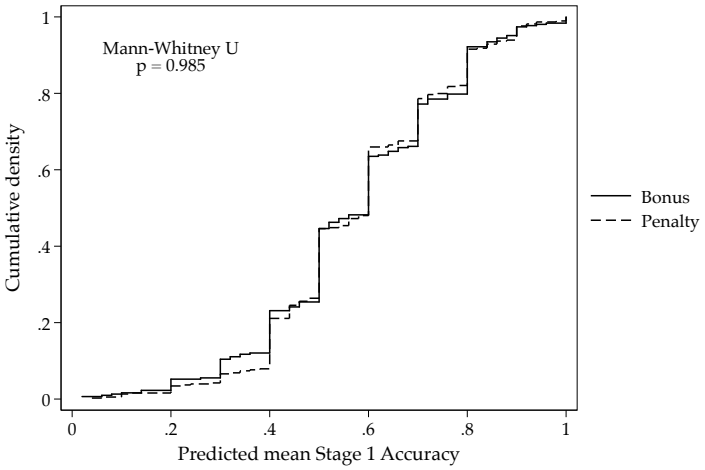


Figure B4: Stage 2 acceptors' predictions of mean stage 1 accuracy.

Table B7: Workers' Predictions of stage 1 Accuracy

	(1) Predicted Acc.	(2) Predicted Acc.
Penalty Frame	0.002	0.001
	(0.015)	(0.014)
High Fixed Pay	0.006	0.010
	(0.020)	(0.019)
High Variable Pay	0.033	0.037*
	(0.020)	(0.020)
Experiment 2	-0.040*	-0.035
	(0.022)	(0.022)
Set dummies	Yes	Yes
Controls	No	Yes
N	686	686
R-squared	0.022	0.151
Mean of dependent variable	0.577	0.577

Dependent variable is worker's prediction of mean stage 1 accuracy. Standard errors clustered at zipcode-experiment level in parentheses. * p<0.10, ** p<0.05, *** p<0.01.
"Set dummies" indicate which set of strings workers typed in stage 1. "Controls" are the full set of regressors from the main specifications.

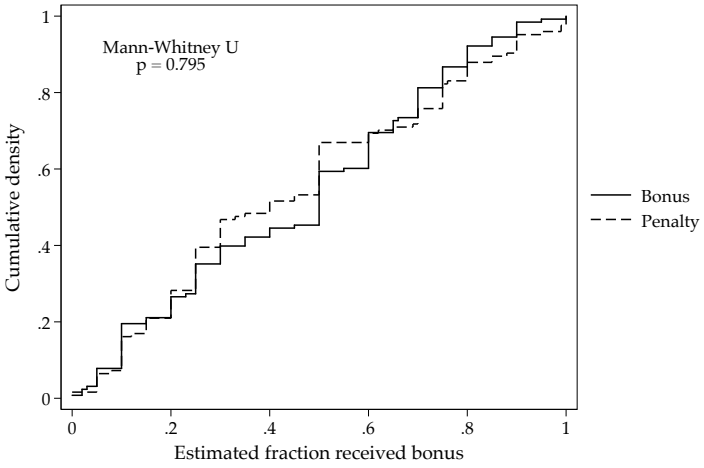


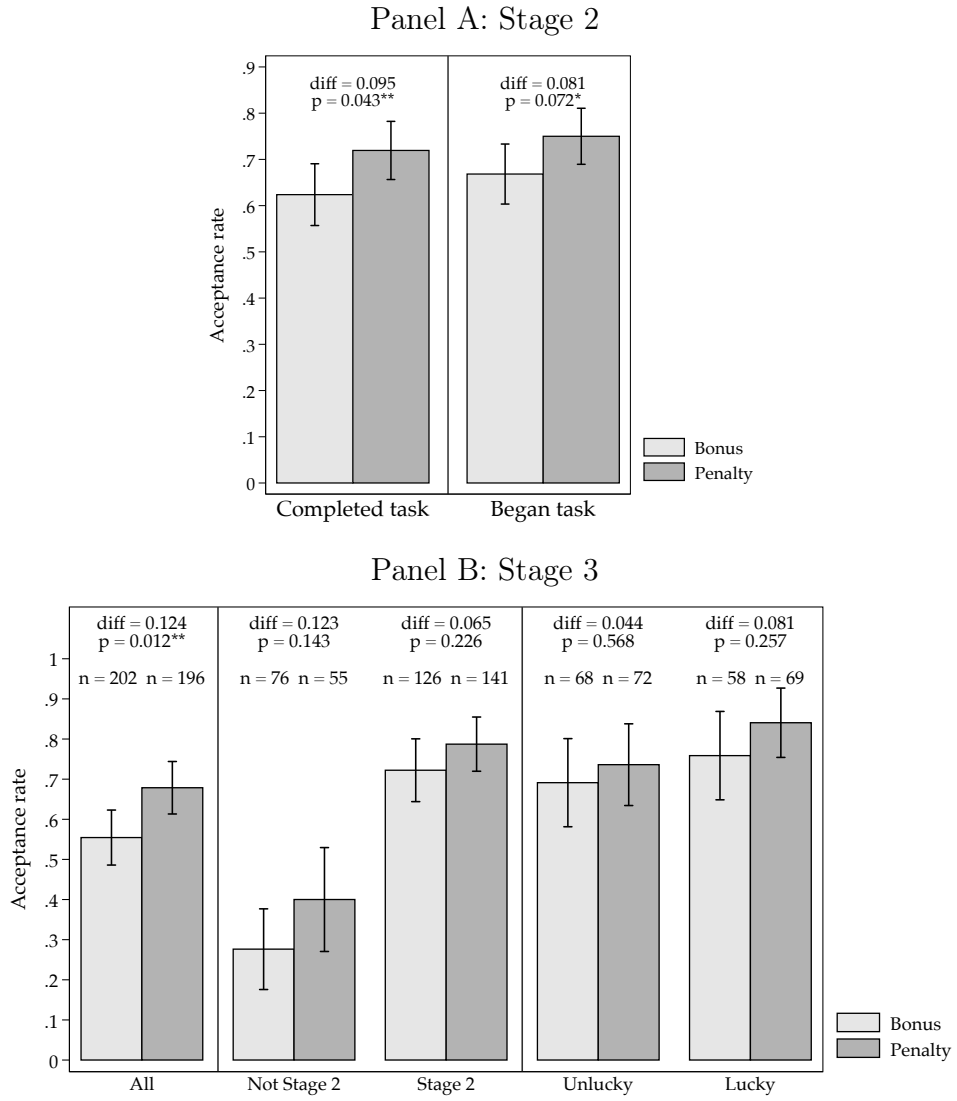
Figure B5: Survey responses to "what proportion of participants do you think received the maximum pay of $3.50."

B.8 Experiment 3 results

Table B8: Selection, Experiment 3

	(1) Accuracy Task 1	(2) Rej. lotteries	(3) Res. wage, $/hr	(4) Fair wage, $/hr
Penalty Frame	0.011	-0.044	0.074	-0.125
	(0.007)	(0.122)	(0.268)	(0.258)
N	267	267	267	267
R-squared	0.008	0.000	0.000	0.001
Mean dep. variable	0.504	0.001	5.257	6.690

Results from regression of key observables on contract terms, conditional on task completion. Dependent variable is indicated above each column. Standard errors clustered at zipcode level in parentheses. * p<0.10, ** p<0.05, *** p<0.01. "Rejected lotteries" and "Accuracy Task 1" are standardized in the full sample to mean zero, standard deviation one.



Notes: First part of panel A shows fraction of workers who completed the stage 2 task, second part shows fraction who began the task but did not necessarily complete. First part of panel B shows fraction of workers who completed stage 3. Second part compares stage 3 completion rates between those who did or did not complete stage 2. Third part compares stage 3 completion rates between those who did or did not correctly guess the paid coin toss in stage 2.

Figure B6: Acceptance of stage 2 and 3 job offers, experiment 3 (coin toss)

B.8.1 Experiment 4: joint evaluation of contract and outside option

I recruited a new sample of 206 MTurk workers, and elicited their valuations of two hypothetical contracts, side-by-side. On the left side was a framed task, while the on the right, a safe and neutrally framed task, representing a safe outside option. Workers were randomly shown either a bonus contract (n=96)

or a penalty contract (n=110) on the left side. The right side was the same for all subjects. The hypothesis was that if considering the framed contract has a "spillover" effect on the valuation of the outside option, I should observe differences in those valuations depending on the framed contract the worker was exposed to.

Subjects were paid $0.20 for a "Short typing task and 2 questions" estimated to take 2 minutes. They first typed five text strings, then were asked to contemplate the two *hypothetical* typing contracts side-by-side, and to report their valuation of each. No other data were collected.

Monetary valuations are potentially difficult to interpret if the framing treatment alters the subject's reference point with respect to money. I therefore elicited valuations in a different dimension: units of effort (time), holding the probability of success constant. The hypothetical contracts were described as requiring X minutes of work with an exogenous success probability of 65%. Subjects reported the maximum X at which they would be willing to accept. Bonus and penalty framed contracts had a fixed pay of $2, variable pay of $2. Safe contracts paid $3 for sure. The questions are reproduced in Web Appendix C.10.

Suppressing w and b, denote the valuation of the framed and safe contracts by $U(0), \bar{u}(0)$ under the bonus treatment and $U(1), \bar{u}(1)$ under the penalty treatment. The average valuations under the bonus treatment were 17.4 minutes (s.d. 11.4) for the bonus contract and 20.4 minutes (s.d. 12.0) for the safe contract. Under the penalty treatment they were 20.1 minutes (s.d. 12.5) for the penalty contract and 20.6 minutes (s.d. 12.6) for the safe contract.[5] Hence $\bar{u}(1) - \bar{u}(0) = 0.26$ minutes (Mann-Whitney U-test p = 0.928, t-test p = 0.882). The framing does *not* appear to influence the valuation of the safe alternative contract. Meanwhile, $U(1) - U(0) = 2.60$ minutes: the penalty contract is valued higher than the bonus contract, though the difference is only significant at 10% in a non-parametric test (Mann-Whitney U-test p = 0.073, t-test p = 0.122). The CDFs in Web Appendix Figure B7 reveal a first-order shift in the distribution of $U(F)$, but not $\bar{u}(F)$.

Turning to the implied preferences between contracts, we can check whether the penalty preference replicates in the new evaluation mode. 17% of bonus subjects and 37% of penalty subjects reported a strictly higher valuation of the contract than the outside option (interpreted as "accepting" the contract), and 26% of bonus subjects and 26% of penalty subjects reported equal valuations

[5]There were two extreme outliers reporting 180 minutes each for the framed contracts (the next highest response was 60). I drop these two subjects for the analysis, noting that they are influential for the parametric (t) tests but not the non-parametric (proportions, rank-sum) tests.

of contract and outside option (indifference). Whether measured as a strict or weak preference, this implies a 20 percentage point higher acceptance rate for the penalty contract (two-tailed tests of proportions, p = 0.001 and p = 0.004 respectively).[6]

Experiment 4 does not support the hypothesis outlined in this section: exposure to a framed contract did not appear to influence the evaluation of a safe outside option. It also provides a fourth replication of the main finding, in a setting with simple payoffs and an exogenous success probability.

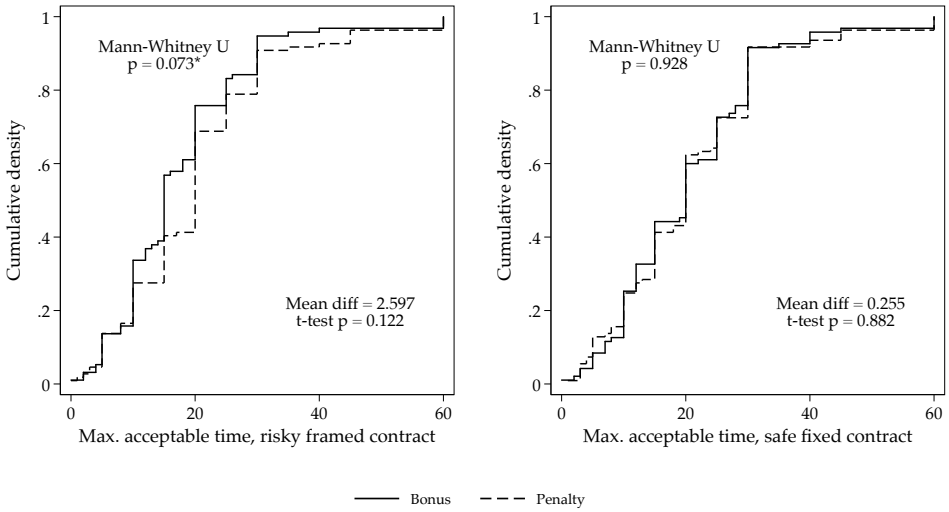


Figure B7: Experiment 4 CDF plots

[6]I can also test whether the relative evaluation of the framed contract is higher under penalties, i.e. a difference-in-differences testing the hypothesis $U(1) - \bar{u}(1) = U(0) - \bar{u}(0)$. A Mann-Whitney U-test rejects (p = 0.003), so does a t-test (p = 0.009).

B.9 Experiment 5 results

Table B9: Balance Check: Experiment 5

	Experiment 5						Exp. 1&2 vs 5	
	Joint		Groups 12 & 13		Groups 14 & 15		Groups 0-7 vs 12-15	
	F-stat	p	Diff	p	Diff	p	Diff	p
Accuracy Task 1	0.65	0.58	-0.01	0.54	0.00	0.94	0.04	0.00***
Hours on Task 1	0.11	0.96	0.01	0.76	-0.00	1.00	0.41	0.00***
Rejected Lotteries	0.20	0.89	0.31	0.26	-0.08	0.79	-0.06	0.61
Inconsistent Lottery Choices	0.60	0.62	0.01	0.60	0.02	0.42	0.00	0.72
Minimum acceptable wage, $/hr	0.25	0.86	0.07	0.77	-0.02	0.93	-0.68	0.00***
Minimum fair wage, $/hr	0.22	0.89	0.08	0.73	0.05	0.88	-0.97	0.00***
Hours working on MTurk per week	1.15	0.33	-0.28	0.83	-1.61	0.21	0.54	0.41
Typical MTurk earnings, $100/week	1.05	0.37	-0.04	0.54	0.01	0.79	-1.42	0.13
MTurk HITs completed	1.03	0.38	-7499.66	0.09*	3609.55	0.07*	-5626.31	0.01***
Months of experience on MTurk	1.17	0.32	1.04	0.49	0.50	0.68	-4.39	0.00***
Mainly participate in research HITs	0.26	0.86	-0.04	0.39	0.04	0.38	-0.04	0.02**
Work on MTurk mainly to earn money	0.04	0.99	-0.03	0.31	-0.03	0.25	0.01	0.44
Age in 2013	0.12	0.95	1.38	0.18	-0.80	0.45	1.05	0.03**
Male	2.11	0.10*	0.08	0.12	-0.00	0.93	0.06	0.01***

"Joint" reports the F-statistic and p-value from a joint test of the significance of the set of group dummies in explaining each relevant baseline variable in experiment 5. Columns labeled "Groups 12 & 13" and "Groups 14 & 15" 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, and groups correspond to those in Table 1. The final two columns report difference in means and p-values for t-tests comparing the "high salience" experiments 1&2 with the "low salience" experiment 5 (a positive difference means this value was higher in experiments 1&2). * p<0.10, ** p<0.05, *** p<0.01.

Table B10: Acceptance decision, Experiment 5

	(1) Accepted	(2) Accepted	(3) Accepted	(4) Accepted	(5) Accepted	(6) Accepted
Penalty Frame	-0.038	-0.050	-0.046	-0.111**	-0.048	-0.048
	(0.035)	(0.050)	(0.048)	(0.052)	(0.049)	(0.049)
Table treatment	-0.055	-0.067	-0.077	-0.099*	-0.073	-0.073
	(0.035)	(0.048)	(0.047)	(0.052)	(0.047)	(0.047)
Penalty * table		0.024	0.026	0.081	0.028	0.028
		(0.069)	(0.068)	(0.075)	(0.069)	(0.069)
Accuracy Task 1			0.574***	0.899**	0.988***	0.988***
			(0.096)	(0.437)	(0.379)	(0.379)
Accuracy Task 1 ^2				-0.387	-0.505	-0.505
				(0.467)	(0.421)	(0.421)
Hours on Task 1			0.011	-0.195	-0.035	-0.035
			(0.144)	(0.185)	(0.151)	(0.151)
Rejected Lotteries			0.029	0.023	0.018	0.018
			(0.018)	(0.021)	(0.019)	(0.019)
Minimum acceptable wage, $/hr			0.011	0.003	0.010	0.010
			(0.010)	(0.011)	(0.010)	(0.010)
Minimum fair wage, $/hr			-0.026***	-0.022**	-0.023**	-0.023**
			(0.009)	(0.010)	(0.009)	(0.009)
Inconsistent lotteries					0.012	0.012
					(0.068)	(0.068)
Set dummies	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	No	No	Yes	Yes	Yes
N	778	778	777	656	776	776
R-squared	0.016	0.016	0.073	0.131	0.118	0.118
Mean dep. variable	0.622	0.622	0.623	0.622	0.622	0.622

Dependent variable is a dummy indicating whether the worker accepted the job offer in stage 2. Standard errors clustered at zipcode level in parentheses. * p<0.10, ** p<0.05, *** p<0.01. One workers is missing timing variable and one missing age variable. Column (4) drops workers 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. "Set dummies" indicate the set of strings workers typed in stage 1. "Controls" are the full set of variables collected in the stage 1 survey. "Rejected lotteries" is standardized to mean zero, standard deviation one.

Table B11: Selection, Experiment 5

	(1) Correct T1	(2) Hours on T1	(3) Rej. lotteries	(4) Res. wage	(5) Fair wage
Penalty Frame	-0.022	-0.008	0.007	0.112	-0.028
	(0.015)	(0.011)	(0.084)	(0.296)	(0.298)
Table treatment	0.004	-0.009	-0.065	-0.218	-0.252
	(0.016)	(0.011)	(0.084)	(0.292)	(0.294)
N	484	484	484	484	484
R-squared	0.004	0.002	0.001	0.001	0.001
Mean dep. variable	0.441	0.253	0.073	5.399	6.750

Table regresses key observables on contract terms, conditional on contract acceptance. Standard errors clustered at zipcode level in parentheses. * p<0.10, ** p<0.05, *** p<0.01. "Rejected lotteries" standardized to mean zero, standard deviation one.

Table B12: Performance on stage 2, Experiment 5

	(1) Accuracy	(2) Accuracy	(3) Accuracy	(4) Accuracy	(5) Accuracy	(6) Accuracy
Penalty Frame	-0.023	-0.007	-0.002	-0.006	0.001	0.001
	(0.018)	(0.025)	(0.021)	(0.023)	(0.021)	(0.021)
Table treatment	0.013	0.029	0.022	0.033	0.021	0.021
	(0.018)	(0.024)	(0.020)	(0.022)	(0.020)	(0.020)
Penalty * table		-0.032	-0.021	-0.033	-0.018	-0.018
		(0.036)	(0.030)	(0.034)	(0.031)	(0.031)
Accuracy Task 1			0.649***	1.194***	1.151***	1.151***
			(0.049)	(0.240)	(0.209)	(0.209)
Accuracy Task 1 ^2				-0.555**	-0.544**	-0.544**
				(0.243)	(0.217)	(0.217)
Hours on Task 1			-0.141*	-0.034	-0.053	-0.053
			(0.074)	(0.093)	(0.078)	(0.078)
Rejected Lotteries			0.002	0.005	0.004	0.004
			(0.009)	(0.010)	(0.009)	(0.009)
Minimum acceptable wage, $/hr			-0.003	-0.006	-0.004	-0.004
			(0.004)	(0.005)	(0.004)	(0.004)
Minimum fair wage, $/hr			-0.001	-0.002	0.000	0.000
			(0.004)	(0.005)	(0.004)	(0.004)
Inconsistent lotteries					-0.025	-0.025
					(0.032)	(0.032)
Set dummies	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	No	No	Yes	Yes	Yes
N	484	484	484	408	483	483
R-squared	0.076	0.077	0.374	0.442	0.422	0.422
Mean dep. variable	0.581	0.581	0.581	0.586	0.581	0.581

Dependent variable is accuracy in the stage 2 typing task, measured as the fraction of items entered correctly. Standard errors clustered at zipcode level in parentheses. * p<0.10, ** p<0.05, *** p<0.01. Column (4) drops workers 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 experiment. "Set dummies" indicate which set of strings workers typed in stages 1 and 2. "Controls" are the full set of variables collected in the stage 1 survey. "Rejected lotteries" is standardized to mean zero, standard deviation one.

B.10 Experiment 6 results

Table B13: Effort in Direct Choice experiment

	(1) Accuracy	(2) Accuracy	(3) Accuracy	(4) Accuracy
Chose penalty	0.004	0.004		
	(0.031)	(0.029)		
Assigned penalty			0.051	0.043
			(0.041)	(0.038)
Accuracy Task 1		0.429***		0.589***
		(0.083)		(0.126)
Strict pref.		-0.014		-0.031
		(0.028)		(0.036)
Order effect		0.037		0.006
		(0.028)		(0.038)
N	168	168	127	127
R-squared	0.000	0.206	0.012	0.258
Mean dep. variable	0.655	0.655	0.657	0.657

Columns (1) and (2) include only workers that chose bonus or penalty. Columns (3) and (4) include those who chose "indifferent." "Strict pref." indicates the "strict preference" treatment. "Order effect" is a dummy equaling one when the bonus contract was shown on the left side. Robust standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

B.11 Expected earnings

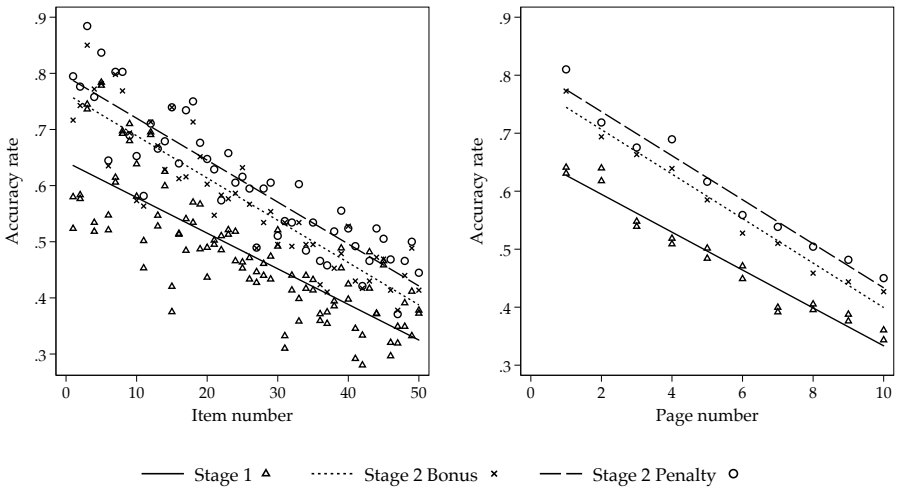
Table B14: Surplus, experiments 1 and 2

	(1) Accepted	(2) Expected Pay	(3) Surplus	(4) Opp. cost	(5) Hours T2	(6) Res. wage
Penalty Frame	0.109***	0.064*	-0.271	0.335*	0.033	0.271
	(0.026)	(0.036)	(0.174)	(0.171)	(0.028)	(0.191)
High Fixed Pay	0.163***	1.534***	1.236***	0.298	0.001	0.370
	(0.036)	(0.030)	(0.243)	(0.241)	(0.038)	(0.258)
High Variable Pay	0.027	0.911***	0.824***	0.087	0.006	0.183
	(0.036)	(0.050)	(0.246)	(0.243)	(0.041)	(0.273)
N	1450	679	679	679	679	679
R-squared	0.031	0.599	0.045	0.009	0.002	0.010
Mean dep. variable	0.474	2.259	-0.812	3.072	0.690	4.555

Expected pay is $w + e * b$ where e is realized stage 2 performance. Opportunity cost is time spent on stage 2 multiplied by reservation wage. Surplus is expected pay minus opportunity cost. Columns (2) to (6) drop 9 workers below the 1st percentile of surplus (each was an outlier for either time spent or reservation wage). Standard errors clustered at zipcode-experiment level in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

B.12 Persistence

Figure B8 plots performance by typed item, or page of five items in stage 1, and separately for each framing treatment in stage 2. Only workers who accepted the stage 2 offer are included. The lines slope down because the text items grow longer between pages. The graph clearly illustrates the shift in performance in the penalty over the bonus frame is persistent throughout the task, there is no evidence of convergence, as confirmed by the regression in Table B15, 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.



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 B8: Performance by item/page on the effort task.

Table B15: 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-experiment level in parentheses. * p<0.10, ** p<0.05, *** p<0.01. "Set dummies" indicate sets of strings workers 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.

B.13 Standard Selection and Incentive Effect

In this section I briefly show that higher ability workers are more likely to accept the stage 2 offers of incentive pay and workers improve their performance between stage 1 (flat pay) and stage 2 (incentive pay).

Performance improved from stage 1 to 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 B9 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. 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 Web Appendix Table B1.

One approach is to assume that the true performance model is linear and equal to that estimated in Table 4 column (4) (excluding the "set" dummy variables for the stage 2 task). Using this model I impute task 2 accuracy for rejectors. The

results are as follows. Mean accuracy across all workers 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 a 13 percentage point improvement. However, the mean fitted stage 2 accuracy for all workers, including rejectors, is 0.56, suggesting 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.

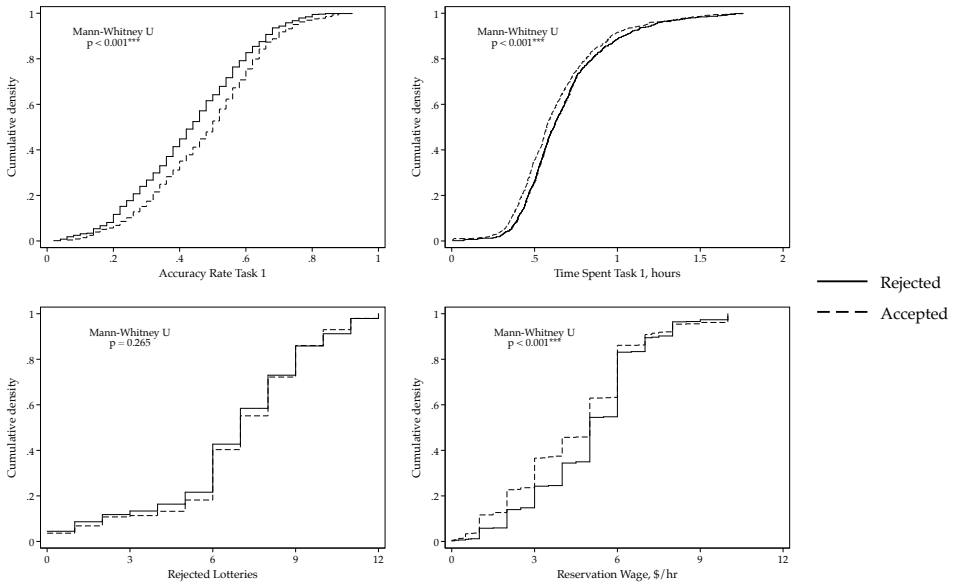


Figure B9: Comparing Acceptors vs Rejectors. Reservation wage trimmed at the 99th percentile.

C Experimental details

C.1 Design timeline

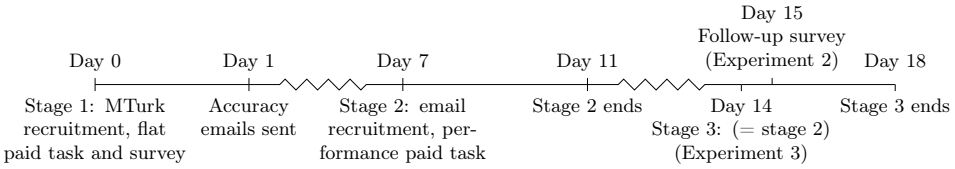


Figure C1: Experiment timeline.

C.2 Informed consent

After clicking the link on MTurk, but before beginning stage 1 of the experiment, workers were required to read and agree to an informed consent statement, which is reproduced below (the form for the coin toss experiment is of course slightly different).

> 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.3 Typing task

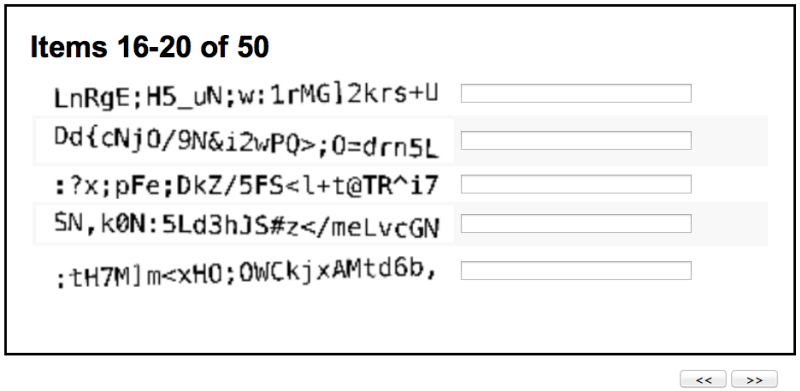


Figure C2: Example screen from the typing task.

C.4 Accuracy report after first stage

Shortly after the first stage was completed, workers 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.5 Invitation to second stage

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

incentive. The following is the full email text from experiment 1. Experiment 2 changed the "pay" section as described in Table 2. Experiment 3 used the same text as experiment 2, but referring to "guesses" rather than typing.

```
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=
XXXXXXXXXXXXX

Your MTurk ID (XXXXXXXXXXXXX) 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 descriptive text.

C.6 Invitation to stage 2, low-salience task

Thanks for participating in our recent typing task and survey. You are invited to do another typing task (typing 30 text strings) in the same format as you did in the previous HIT. There is no survey this time.

To learn more about the task and payment, please use the following personalized link.

https://qeurope.eu.qualtrics.com/SE/?SID=SV_eKAUmLmw1akCrsN&MID=XXXXXXXXXXXXXX

Your MTurk ID (XXXXXXXXXXXXXX) will be recorded automatically.

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

C.7 Stage 2, low-salience task

Conventional framing:

Typing Task
Please read the following carefully

This task involves typing 30 text strings, in the same format as you did in the previous HIT. There is no survey this time.

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 $2 which will be reduced by $1 if the checked item

```
is incorrect.

Payment:
Your MTurk ID has been recorded automatically, and we'll confirm
it at the end of the task.  Payments will be made through the
MTurk "bonus system" within 48 hours of the task closing.

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

Table format:

```
Typing Task
Please read the following carefully

This task involves typing 30 text strings, in the same format as
you did in the previous HIT. There is no survey this time.

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.

 Total pay if correct:    $2
 Total pay if incorrect:  $1

Payment:
Your MTurk ID has been recorded automatically, and we'll confirm
it at the end of the task.  Payments will be made through the
MTurk "bonus system" within 48 hours of the task closing.

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

C.8 Direct choice (Experiment 6)

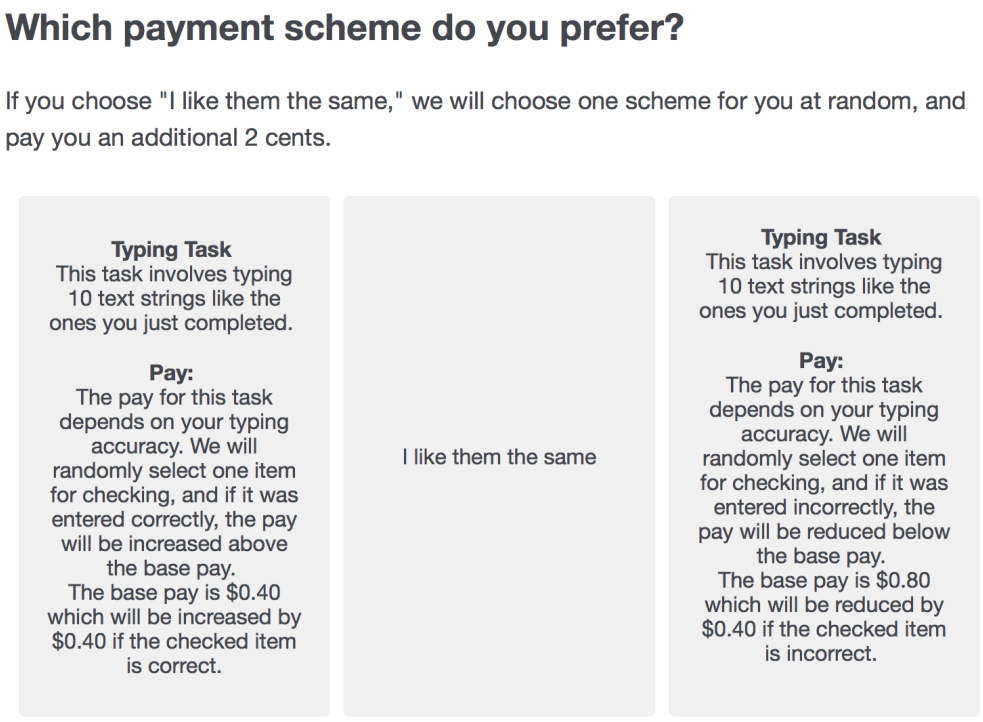


Figure C3: Screenshot from direct choice task.

C.9 Lottery Questions

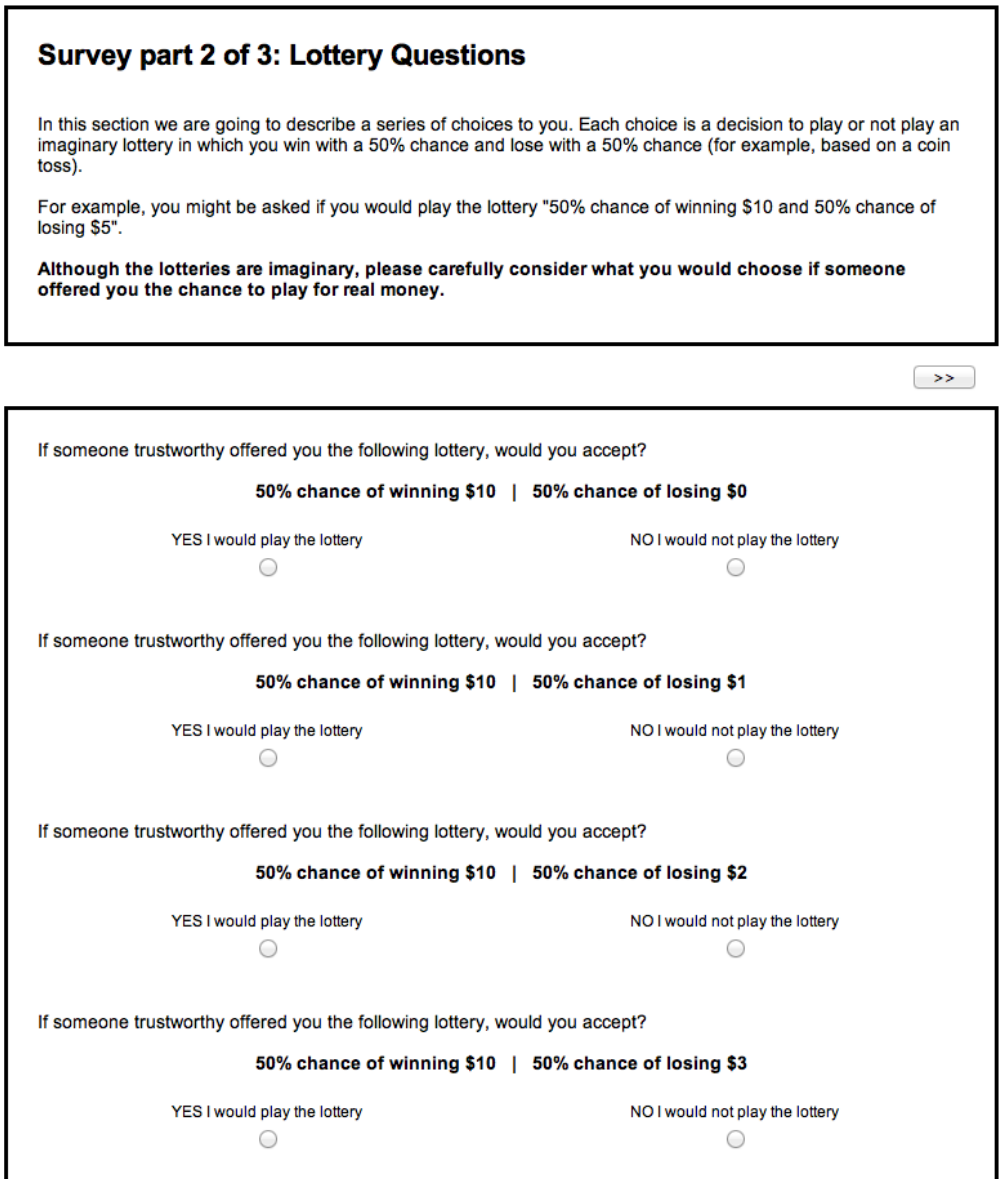
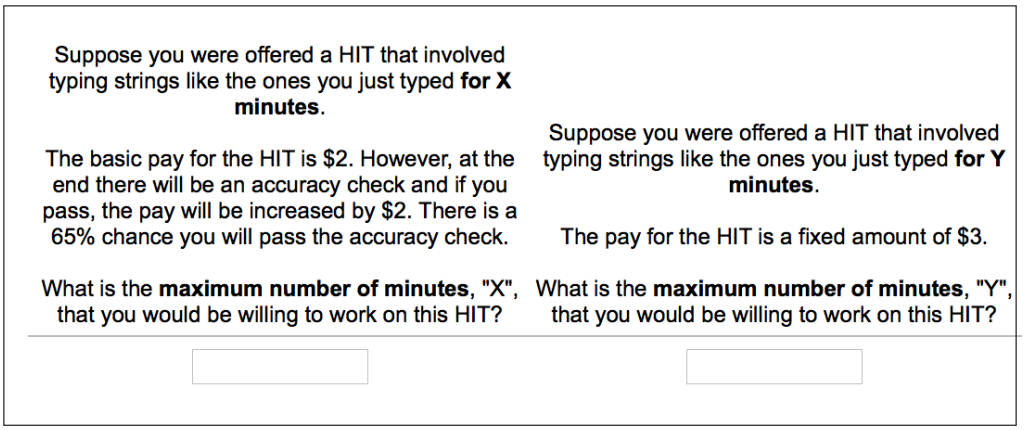


Figure C4: Introduction and examples of lottery questions.

C.10 Experiment 4 Questions

Bonus treatment



Penalty treatment

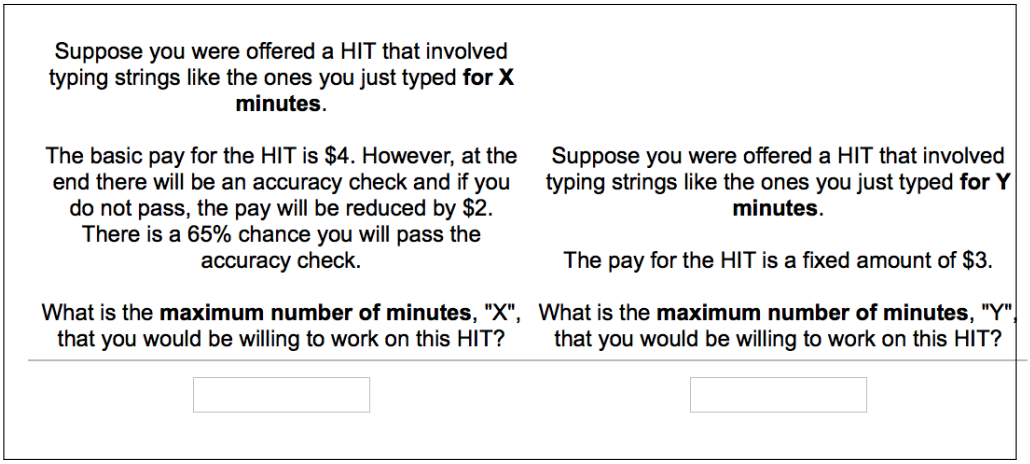


Figure C5: Experiment 4 questions

References

Crawford, V. P. and J. Meng (2011). 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.

Gill, D. and V. Prowse (2012). 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: Explain-

ing Ambitious Goal-setting and Nonconvex Preferences. *mimeo*.

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

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

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

Koch, A. K., J. Nafziger, A. Suvorov, and J. van de Ven (2014). Self-Rewards and Personal Motivation. *European Economic Review 68*, 151–167.

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

Tversky, A. and D. Kahneman (1981). The framing of decisions and the psychology of choice. *Science 211*(4481), 453–458.