Bonus versus Penalty: How Robust Are the Effects of Contract Framing?

24 August 2017

Jonathan de Quidt[a], Francesco Fallucchi[b], Felix Kölle[c], Daniele Nosenzo[d] and Simone Quercia[e]

Abstract:
We study the relative effectiveness of contracts that are framed either in terms of bonuses or penalties. In one set of treatments subjects know at the time of effort provision whether they have achieved the bonus / avoided the penalty. In another set of treatments subjects only learn the success of their performance at the end of the task. We fail to observe a contract framing effect in either condition: effort provision is statistically indistinguishable under bonus and penalty contracts.

Keywords: contract framing; bonus; penalty; fine; loss aversion.

JEL Classification Numbers: C9; D03; J24

Acknowledgements: We thank the Editor, Robert Slonim, and two anonymous reviewers for helpful comments. We also received useful suggestions from seminar participants in Bonn, Nottingham, and Warwick. This work was supported by the Economic and Social Research Council [grant number ES/K002201/1]. de Quidt acknowledges financial support from Handelsbanken's Research Foundations, grant no: B2014-0460:1. We thank Benjamin Beranek and Lucas Molleman for their help in implementing the experiments.

[a] Institute for International Economic Studies, Stockholm University.
[b] Luxembourg Institute of Socio-Economic Research.
[c] Center for Social and Economic Behavior, University of Cologne.
[d] School of Economics, University of Nottingham.
[e] Institute for Applied Microeconomics, University of Bonn.

1. INTRODUCTION

Although incentive pay can be very effective in raising employees' performance (e.g., Lazear, 2000), the way incentives are described also matters. A recent experimental literature suggests that incentives are more effective when they are framed as penalties for poor performance rather than bonuses for good performance. For example, Hannan et al. (2005) found that employees exerted significantly more effort under a "penalty" contract that paid a base salary of $30 minus a $10 penalty if they did not meet a performance target, than under a "bonus" contract that paid $20 plus a bonus of $10 if the target was met. The two contracts are isomorphic and so the increase in effort is entirely due to a *framing effect* (Tversky and Kahneman, 1981). Several other studies confirmed this finding, both in the lab (Armantier and Boly, 2015; Imas et al., 2017) and in the field (Fryer et al., 2012; Hossain and List, 2012; Hong et al., 2015).

The size of this framing effect is large. Figure 1 (left panel) shows effect sizes and confidence intervals of the three lab experiments cited above (Hannan et al., 2005; Armantier and Boly, 2015; Imas et al., 2017).[1] The average Hedges' g statistic across these studies is 0.51 (Hedges, 1981).[2] However, Figure 1 (right panel) also shows that three further studies found considerably smaller effects, which are either statistically insignificant (DellaVigna and Pope, 2016; Grolleau et al., 2016), or only marginally significant (Brooks et al., 2012).

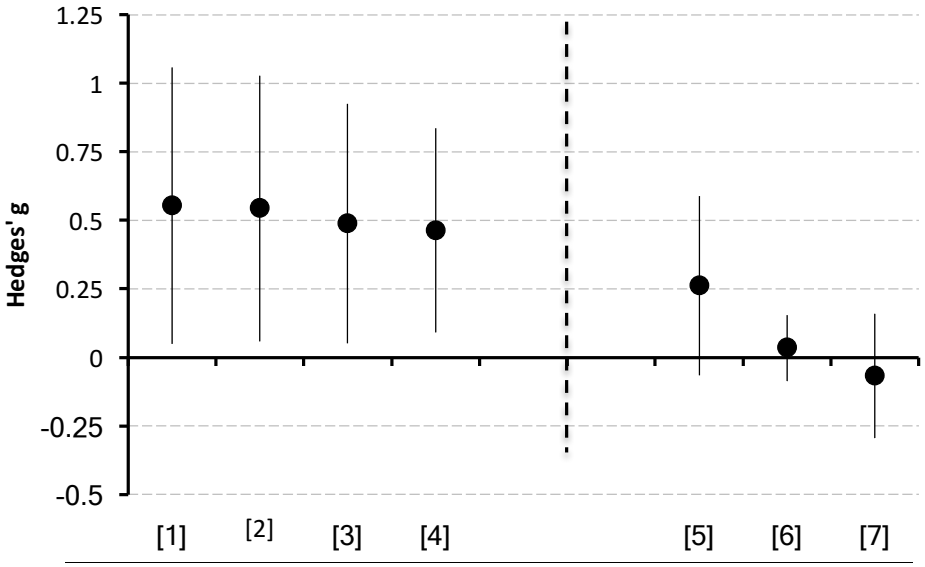
One systematic difference between the experiments in the left and right panels of Figure 1 relates to whether or not subjects could check during the task whether they had met the performance target and hence their monetary compensation. In Brooks et al. (2012), DellaVigna and Pope (2016) and Grolleau et al. (2016), subjects were told in advance what the target was and could verify their monetary compensation at any point during the

[1] Armantier and Boly (2015) ran one experiment in Burkina Faso and one in Canada, and therefore we report two effect sizes. We do not include de Quidt (2017) in Figure 1 because his subjects were allowed to self-select into treatment.

[2] Hedges' g is similar to Cohen's d with a correction for small samples. It is measured as $g = (\mu_p - \mu_b)/\sqrt{[(n_p - 1)sd_p^2 + (n_b - 1)sd_b^2]/(n_p + n_b - 2)}$ where μ_p and μ_b are the mean efforts in the penalty and bonus treatments, sd_p and sd_b their respective standard deviation and n_p and n_b the sample sizes. Thus, Hedges' g measures the standardized difference in effort due to the framing of the incentive scheme. The studies shown in Figure 1 vary, however, in the level of incentives (i.e. the ratio between base salary and bonus/penalty) offered to subjects, and one should keep this in mind when comparing effect sizes across Figure 1 since the effect of framing might vary with the incentive level (Armantier and Boly, 2015).

experiment.[3] This is not the case for the other studies in Figure 1.[4] However, there are many other differences across these studies, which makes it difficult to draw definitive conclusions about the exact causes of the discrepancy in effect sizes (Table A1 in Appendix A summarizes the characteristics of the studies included in Figure 1). In this paper, we report an experiment designed to replicate the pattern displayed in Figure 1 by testing whether the effectiveness of contract framing depends on the availability of information about the performance target.

Figure 1 – Effect size of contract framing in previous experiments



Note: Effect sizes are computed using Hedges' g (Hedges, 1981). Bars represent 95% confidence intervals computed as $g \pm 1.96 * \left\{\sqrt{\left[\frac{n_p+n_b}{n_p n_b} + \frac{g^2}{2(n_p+n_b)}\right]}\right\}$. [1] = Armantier/Boly (2015) - Burkina Faso; [2] = Hannan et al (2005); [3] = Imas et al. (2017); [4] = Armantier/Boly (2015) - Canada; [5] = Brooks et al. (2012); [6] = DellaVigna and Pope (2016); [7] = Grolleau et al. (2016).

We describe our experiment design in Section 2. Subjects performed a real-effort task under either a bonus or penalty contract. Both contracts specified a base pay and an extra

[3] In DellaVigna and Pope (2016) subjects participated in a real-effort task and received a bonus/penalty based on a performance target that was specified ex-ante. Brooks et al. (2012) conducted a chosen-effort experiment where subjects knew in advance whether any level of effort resulted in a bonus/penalty. Grolleau et al. (2016) gave subjects pairs of matrices containing numbers and, in each pair, subjects had to find two numbers that added up to ten. They were paid a piece-rate for each correctly-solved pair of matrices.

[4] In Imas et al. (2017) incentives were contingent upon meeting a target that was not specified ex-ante. In Hannan et al. (2005) effort only affected the probability of meeting the target. In Armantier and Boly (2015) participants were recruited to spell-check exam papers and bonuses/penalties depended on the quality of their spell-checking (verified ex-post by the experimenters).

amount of money that subjects could earn by meeting a performance target. In the bonus contract, subjects were told that they could increase their base pay by meeting the target. In the penalty contract, they were told that the base pay would be reduced if they missed the target. We implemented these contracts under two conditions. In one condition, akin to the studies in the left panel of Figure 1, the performance target was not specified ex-ante: subjects were told that their performance would be compared with the average performance in a previous experiment. In the other condition we announced the target at the beginning of the task, as in the studies in the right panel of Figure 1.

We report our results in Section 3. Performance in the real-effort task is statistically indistinguishable under the bonus and penalty contracts, both under announced and unannounced targets. While the absence of contract framing effects under announced targets is consistent with the existing evidence, our experiment fails to replicate the findings of the studies displayed on the left panel of Figure 1 that had found significant framing effects when the target was unannounced. We discuss the implications of these results in Section 4.

2. Experimental Design

Our experiment was conducted online with 853 subjects recruited on Amazon's Mechanical Turk (MTurk).[5] The experiment consisted of 3 parts plus a questionnaire. Subjects knew this in advance, although they did not receive instructions for each part until they had completed the previous ones. Only one part, randomly selected at the end, was paid out.

In Part 1, subjects participated in the "Encryption Task" (Erkal et al., 2011): they had to encode a series of words by substituting letters with numbers using predetermined letter-to-number assignments. Subjects had 5 minutes to encode as many words as possible and were paid $0.05 per word, while receiving live feedback about the total number of words encoded so far.[6] This part of the experiment was the same across treatments and is used to obtain a baseline measurement of subjects' ability in the task.

Part 2 varied across treatments according to a 2x2 between-subject design. In all treatments subjects had again to encode words and were paid based on how many words they

[5] MTurk is an online labor market. The experiments were conducted using the software LIONESS (Aréchar et al., 2017). Subjects were adult residents of the US. See Horton et al. (2011) for a discussion of experiments conducted on MTurk.

[6] Subjects had to encode a word correctly before they could proceed to the next. The letter-to-number assignments were kept constant across the whole experiment. See Appendix B for the experimental instructions.

encoded within 10 minutes, again with live feedback on the total number of words encoded. In the **Bonus** treatments the payment specified a base pay of $0.50 plus a bonus of $1.50 if the subject encoded as many words as specified in a productivity target. In the **Penalty** treatments the payment specified a base pay of $2.00 minus a penalty of $1.50 if the target was not met. In the **Announced** treatments the target was announced at the beginning of the task.[7] In the **Unannounced** treatments the target was left unspecified: subjects were just told that the target was based on the average productivity of participants in a previous study.[8]

Part 3 was again the same in all treatments. One of the explanation for the existence of contract framing effects suggested in the literature is loss aversion. In order to assess the role of loss aversion in our experiment, we used the lottery choice task introduced by Gächter et al. (2010). Subjects received a list of 6 lotteries and decided, for each lottery, whether to accept it (and receive its realization as a payment) or reject it (and receive nothing). Each lottery specified a 50% probability of winning $1.00 and a 50% probability of losing an amount of money that varied across lotteries from $0.20 to $1.20, in $0.20 increments.[9] As discussed by Gächter et al. (2010), a subject's pattern of acceptances/rejections in this task measures his/her degree of loss aversion.

Finally, subjects completed a short questionnaire measuring standard socio-demographics. Additionally in the Unannounced treatments subjects were asked to guess what the target was before learning the outcome of the experiment.[10]

Table 1 summarizes the design.[11] Sample sizes were determined using power analysis. Based on the average effect size ($g = 0.51$) reported across the three studies with

---

[7] We set the target at 45 words based on the results of a pilot conducted to calibrate incentives.

[8] This is similar to Imas et al. (2017). The target was based on the average performance of subjects in the Announced condition and equal to 39 words. Note that this might induce an additional effort motive in Unannounced: because there is an implicit comparison to the performance of others, some workers may respond differentially to penalties because they expect that *other* workers will. Also note that the target is lower in Unannounced than in Announced. Because the true value of the target was announced only at the end of the experiment, this should not affect performance.

[9] In Part 3 subjects were initially given $1.20 and losses were subtracted from this initial payment. At the end of the experiment, if Part 3 was selected for payment, one lottery was chosen at random and, if accepted, played out to determine final payments.

[10] This belief elicitation was not incentivized. The average belief was 32 in Bonus (s.d. = 12.8) and 31 in Penalty (s.d. = 12.1), and the difference is insignificant ($p = 0.899$ using a two-sided Mann-Whitney test; $p = 0.769$ using a two-sided Kolmogorov-Smirnov test; 137 observations per treatment).

[11] We conducted one additional treatment (N = 140) where in Part 2 subjects received a flat payment of $0.50, regardless of how many words they encoded. With this control treatment we assess whether subjects are responsive to the type of incentives used in the Bonus/Penalty treatments. (We thank an anonymous referee for suggesting this treatment). This could be important since MTurkers may be partly motivated by reputation (e.g., their work needs to be "approved" by the employer before a payment can be made and the approval rate is a statistic that future employers can use to screen workers). Reputational concerns may dilute the effect of short-

unannounced targets, we assigned 137 observations to each of the Bonus and Penalty treatments in the Unannounced condition. This gives us 98% power to detect the original average effect size at the 5% level of significance. We assigned our remaining resources to recruit subjects in the Announced condition. Given the resulting sample size (292 subjects in Bonus; 287 in Penalty), we have an 80% power to detect an effect size of at least 0.24 at a 5% level of significance.[12]

Table 1 – Experiment design

	Unannounced treatment	Announced treatment
Bonus treatment	Contract pays **$0.50 + $1.50** if subject meets an **unspecified** performance target *(N = 137)*	Contract pays **$0.50 + $1.50** if subject meets a **pre-specified** performance target *(N = 292)*
Penalty treatment	Contract pays **$2.00 - $1.50** if subject does not meet an **unspecified** performance target *(N = 137)*	Contract pays **$2.00 - $1.50** if subject does not meet a **pre-specified** performance target *(N = 287)*

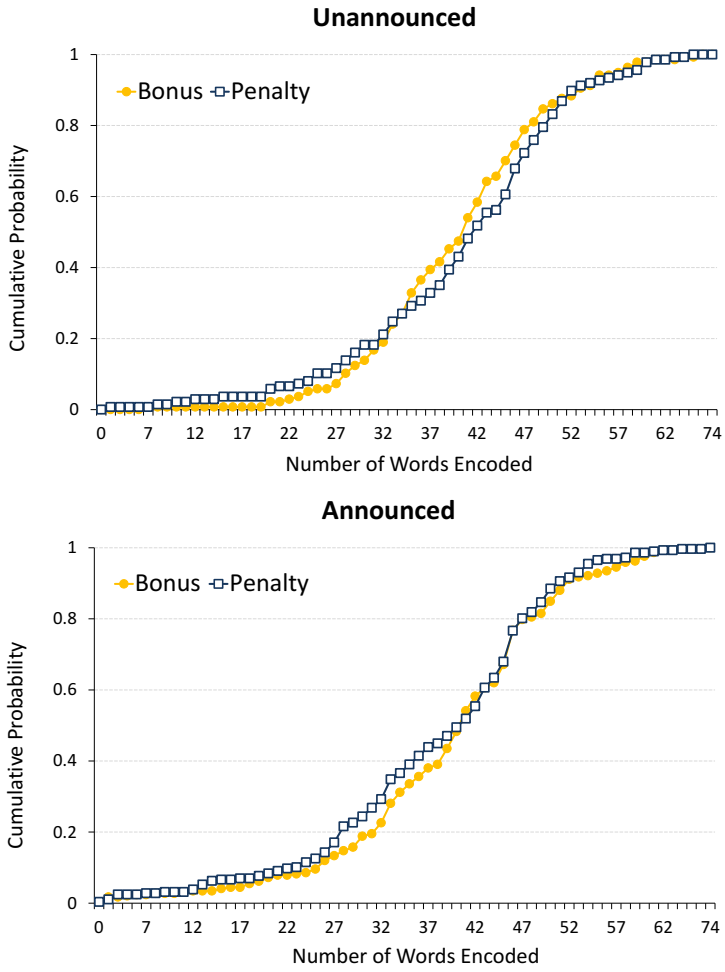
3. RESULTS

Figure 2 shows the cumulative distribution functions (CDFs) of the numbers of words encoded by participants in Part 2 of the experiment. The top and bottom panels show the CDFs of the Bonus and Penalty treatments for the Unannounced and Announced conditions, respectively.

term monetary incentives offered in the experiment, making it harder to find framing effects. We find large differences in effort between the control and Bonus/Penalty treatments (see Appendix C), suggesting that MTurkers are responsive to the type of monetary incentives offered in our task. See also DellaVigna and Pope (2016) fur further corroborating evidence.

[12] Given the average effect size of 0.08 reported across the three studies with announced targets, the achieved power in the Announced condition is low (16%). An 80% power could only be achieved with a sample of about 2450 subjects per treatment. This is due to the very small effect sizes reported in DellaVigna and Pope (-0.07) and Grolleau et al. (0.03). The effect size reported in Brooks et al. (0.26) can be detected given our sample size.

Figure 2 – Performance across treatments



First of all, note that in Announced we observe about 23% of subjects encoding more than the target of 45 words. There are three possible explanations for this result: one is that, in addition to extrinsic incentives, workers are intrinsically motivated to provide effort, perhaps because they enjoy the task. Another possibility is that subjects care about reputation on top of monetary incentives. A third possibility is a gift-exchange hypothesis: since workers are always being paid something, they respond by providing effort.[13]

Regarding contract framing effects, in both conditions the CDFs of Bonus and Penalty overlap substantially, indicating very small differences in performance. In Unannounced, subjects encoded on average 41 words (s.d. = 11.5) in Penalty and 40 words (s.d. = 9.78) in Bonus. The difference is statistically insignificant ($p = 0.407$ using a two-sided Mann-Whitney test; $p = 0.513$ using a two-sided Kolmogorov-Smirnov test; 137 observations per

[13] However, as shown by our control experiment reported in Appendix C, all these additional motives do not seem to offset the effectiveness of monetary incentives.

treatment). In Announced, subjects in Penalty encoded on average fewer words (38; s.d. = 12.4) than in Bonus (39; s.d. = 12.0). This difference is also insignificant ($p = 0.291$ using a two-sided Mann-Whitney test; $p = 0.383$ using a two-sided Kolmogorov-Smirnov test; 292 and 287 observations in Bonus and Penalty, respectively). Moreover, we find no difference between contract framing effects between Announced and Unannounced conditions.[14]

4. Discussion and Conclusion

In our experiment subjects perform a real-effort task and are paid for meeting a performance target. The incentives are framed either as "bonuses" or "penalties" for meeting / not meeting the target. We conducted two sets of experiments in which the target was either announced at the beginning of the task or not. In both settings we find that performance is statistically indistinguishable between bonus and penalty frames. The absence of a contract framing effect when the target is announced is consistent with the findings of Brooks et al. (2012), DellaVigna and Pope (2016) and Grolleau et al. (2016). However, our finding that performance is unresponsive to framing when the target is unannounced contrasts with results reported by Hannan et al. (2005), Armantier and Boly (2015) and Imas et al. (2017).

What can explain our failure to replicate a contract framing effect when the target is unannounced? First of all, we emphasize that, given the average effect size observed in the literature (about 0.5), our study is highly powered, and so the null result is not due to a lack of power to detect an effect of such size. Thus, one way to interpret our results is that there is no effect of framing on effort provision. However, this conclusion is conditional on the true magnitude of the effect being as large as reported in previous studies. Moreover, this leaves unexplained why several previous studies did find significant framing effects.

We believe that a more plausible interpretation of our results is that the "true" effect of contract framing is simply smaller than previously reported. We can conduct a meta-analysis of the effect sizes observed in the literature to get a more precise estimate of the true effect. Using the effect sizes and standard errors reported by Hannan et al., Armantier and Boly and Imas et al., as well as our own Unannounced treatments, we can compute a weighted mean

[14] See Appendix D where we report a pooled regression with dummies for Penalty, Unannounced, and their interaction. The coefficient of the interaction term is very small (0.13) and insignificant ($p = 0.922$). In Appendix D we also report an analysis of the interaction of framing effects and degree of loss aversion, as measured in the lottery task of Part 3 of the experiment. We find no evidence of a contract framing effect even among the most loss averse participants.

estimate of the effect size equal to 0.313.[15] We can repeat the analysis for the Announced condition, combining our data with that of Brooks et al., DellaVigna and Pope and Grolleau et al.. The weighted mean estimate of the effect size is 0.003. Finally, we can compute an estimate of the effect size combining the two conditions and using the data from all studies reported in Figure 1 as well as our treatments. This is equal to 0.071.

Finally, a word of caution should be spent about the specific subject pool used in our study, MTurk workers. One may worry that the small effect of framing in our study is due to the fact that MTurkers are generally unresponsive to the type of (small) monetary incentives used in experiments (e.g., because they mainly care about reputation). However, as discussed above, this is unlikely to be the case: the pay-per-performance incentives used in the experiment raise effort substantially relative to a control treatment with flat payments (see footnote 11 and Appendix C). Nevertheless, there is some evidence that interventions that rely on subtle psychological manipulations, like contract framing, may produce somewhat weak effects in this setting: DellaVigna and Pope (2016), for example, find limited evidence of contract framing effects as well as of probability weighting on a large sample of MTurkers. Similarly, while the experiment conducted by de Quidt (2017) on MTurk identifies a significant contract framing effect, the reported effect size is smaller (about 0.2) than those reported in previous lab experiments.[16] While our study cannot draw definitive conclusions about the role of subject pool idiosyncrasies, this seems an important question for future research.

[15] This is computed as $\sum_{i=1}^{k} w_i g_i / \sum_{i=1}^{k} w_i$ where g_i is the effect size of study i and w_i is its weight, equal to the inverse of the within-study variance.

[16] However, in his study subjects self-selected into treatment, which may affect the estimate of the effect size.

REFERENCES

Aréchar, A.A., L. Molleman, and S. Gächter. 2017. Conducting interactive decision making experiments online. CeDEx Discussion Paper 2017-02.

Armantier, O., and A. Boly. 2015. Framing of Incentives and Effort Provision. *International Economic Review* 56(3), 917–938.

Brooks, R.R.W., A. Stremitzer, and S. Tontrup. 2012. Framing Contracts - Why Loss Framing Increases Effort. *Journal of Institutional and Theoretical Economics* 168(1), 62–82.

DellaVigna, S., and D. Pope. 2016. What Motivates Effort? Evidence and Expert Forecasts. NBER Working Paper No.22193.

Erkal, N., L. Gangadharan, and N. Nikiforakis. 2011. Relative Earnings and Giving in a Real-Effort Experiment. *American Economic Review* 101(7), 3330–48.

Fryer, R.G.J., S.D. Levitt, J.A. List, and S. Sadoff. 2012. Enhancing the Efficacy of Teacher Incentives through Loss Aversion: A Field Experiment. NBER Working Paper No. 18237.

Gächter, S., E.J. Johnson, and A. Herrmann. 2010. Individual-level loss aversion in riskless and risky choices. CeDEx Discussion Paper 2010-20.

Grolleau, G., M.G. Kocher, and A. Sutan. 2016. Cheating and loss aversion: do people lie more to avoid a loss? *Management Science* 62(12), 3428–3438.

Hannan, L.R., V. Hoffman, and D. Moser. 2005. Bonus versus Penalty: Does Contract Frame Affect Employee Effort? In *Experimental Business Research: Economic and Managerial Perspectives*, ed by. Amnon Rapoport and Rami Zwick, 2:151–169. Holland: Springer.

Hedges, L.V. 1981. Distribution Theory for Glass's Estimator of Effect size and Related Estimators. *Journal of Educational and Behavioral Statistics* 6(2), 107–128.

Hong, F., T. Hossain, and J.A. List. 2015. Framing manipulations in contests: A natural field experiment. *Journal of Economic Behavior & Organization* 118, 372–382.

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* 58(12), 2151–2167.

Imas, A., S. Sadoff, and A. Samak. 2017. Do People Anticipate Loss Aversion? *Management Science* 63(5), 1271–1284.

Lazear, E.P. 2000. Performance Pay and Productivity. *American Economic Review* 90(5), 1346–1361.

de Quidt, J. 2017. Your Loss Is My Gain: A Recruitment Experiment With Framed Incentives. *Journal of the European Economic Association, forthcoming*.

Tversky, A., and D. Kahneman. 1981. The Framing of Decisions and the Psychology of Choice. *Science* 211(4481), 453–458.

SUPPLEMENTARY MATERIAL (NOT FOR PUBLICATION)

APPENDIX A: DETAILS OF STUDIES INCLUDED IN FIGURE 1

Table A1 shows details of the experimental designs of previous studies on incentives framing included in Figure 1 in the main text.

Table A1 – List of studies included in Figure 1 in the main text

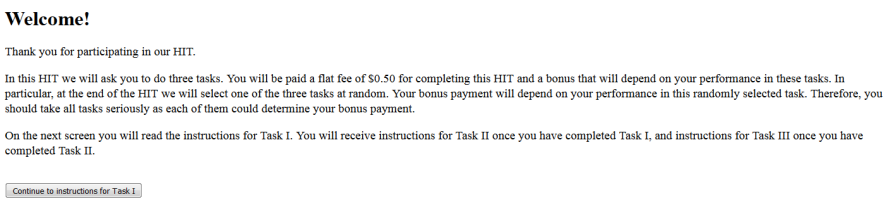
Study	Country	Subject pool	N	Task	Incentive scheme	Target announced?
Armantier & Boly (2015)*	Burkina Faso	Students and general population with a degree	B = 29 P = 34	Grading exams	Base pay: FCFA500. B/P: three levels FCFA 1500/2500/4500, depending on the "grading quality", i.e. the proportion of mistakes correctly identified.	No
Hannan et al. (2005)	United States	M.B.A. students	B = 35 P = 33	Chosen effort	Base pay: $20 B/P: $10	No
Imas et al. (2017)	United States	Students	B = 40 P = 43	Slider task	Base pay: none B/P: t-shirt (cost $9)	No
Armantier & Boly (2015)*	Canada	Students	B = 58 P = 56	Grading exams	Base pay: C$2.08 B/P: three levels, depending on target, C$6.25/10.42/18.75	No
Brooks et al. (2012)	Switzerland	Students	B = 72 P = 73	Chosen effort	Base pay: CHF 20 B/P: CHF 5	Yes
DellaVigna and Pope (2016)	United States	MTurkers	B = 545 P = 532	Pressing a-b keys	Base pay: $1 B/P: $0.4	Yes
Grolleau et al. (2016)	France	Students	B = 150 P = 150	Finding two numbers that add up to 10 in pairs of matrices	Base pay: none B/P: €1.5 per pair of matrices solved/unsolved	Yes

Notes: B = Bonus; P = Penalty; FCFA = CFA Franc; $ = US Dollar; C$ = Canadian Dollar; CHF = Swiss Franc; € = Euro
* Armantier & Boly (2015) also conduct a treatment with both bonus and penalties, depending on performance (N=34 in Burkina Faso and N=56 in Canada);

APPENDIX B: EXPERIMENTAL INSTRUCTIONS

[Common to all treatments]

Introduction



Welcome!

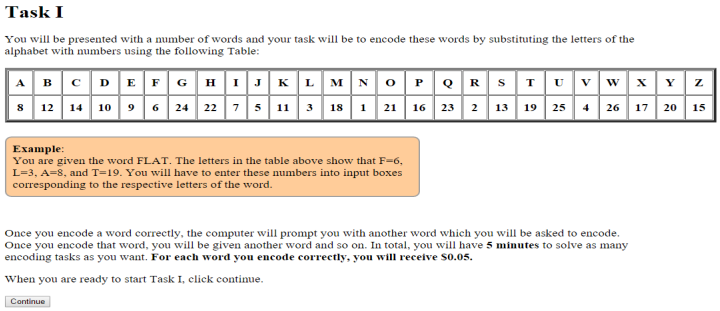
Thank you for participating in our HIT.

In this HIT we will ask you to do three tasks. You will be paid a flat fee of $0.50 for completing this HIT and a bonus that will depend on your performance in these tasks. In particular, at the end of the HIT we will select one of the three tasks at random. Your bonus payment will depend on your performance in this randomly selected task. Therefore, you should take all tasks seriously as each of them could determine your bonus payment.

On the next screen you will read the instructions for Task I. You will receive instructions for Task II once you have completed Task I, and instructions for Task III once you have completed Task II.

Continue to instructions for Task I

Part 1 instructions



Task I

You will be presented with a number of words and your task will be to encode these words by substituting the letters of the alphabet with numbers using the following Table:

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
8	12	14	10	9	6	24	22	7	5	11	3	18	1	21	16	23	2	13	19	25	4	26	17	20	15

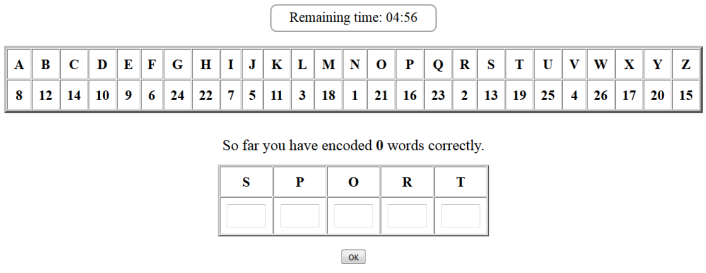
Example:
You are given the word FLAT. The letters in the table above show that F=6, L=3, A=8, and T=19. You will have to enter these numbers into input boxes corresponding to the respective letters of the word.

Once you encode a word correctly, the computer will prompt you with another word which you will be asked to encode. Once you encode that word, you will be given another word and so on. In total, you will have **5 minutes** to solve as many encoding tasks as you want. **For each word you encode correctly, you will receive $0.05.**

When you are ready to start Task I, click continue.

Continue

Part 1 task



Remaining time: 04:56

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
8	12	14	10	9	6	24	22	7	5	11	3	18	1	21	16	23	2	13	19	25	4	26	17	20	15

So far you have encoded **0** words correctly.

S	P	O	R	T

OK

Reminder: you will earn $0.05 for each word you encode correctly.

Part 1 feedback

This is the end of Task I

You have encoded 0 words correctly.
If this task is selected for payment you will receive a bonus earning of $0.00.

On the next screen you will read the instructions for Task II.

Continue to instructions for Task II

[Unannounced Fine]

Part 2 instructions

Task II

In this task you will again encode words using the following Table:

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
8	12	14	10	9	6	24	22	7	5	11	3	18	1	21	16	23	2	13	19	25	4	26	17	20	15

Like in Task I, once you have encoded a word correctly the computer will prompt you with another word to encode. This time, however, you will have **10 minutes** minutes to encode as many words as you want.

For Task II you will receive an initial payment of $2.00. Moreover, other MTurkers participated in this same task in a previous study and earned cash based on their performance. Your individual performance will be compared to the average number of words encoded by these participants. **If you encode a number of words lower than the average of the participants in the previous study, we will reduce your payment by $1.50 (so you will receive $0.50 in total).**

When you are ready to start Task II, click continue.

Continue

Part 2 task

Remaining time: 09:54

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
8	12	14	10	9	6	24	22	7	5	11	3	18	1	21	16	23	2	13	19	25	4	26	17	20	15

So far you have encoded **0** words correctly.

B	E	D

OK

You will receive an initial payment of $2.00. If you encode fewer words than the average of the previous study, this payment reduces to $0.50.

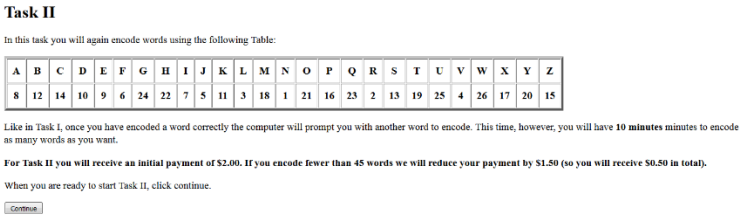
Part 2 belief

Before continuing, we ask you to guess the average number of words encoded by the participants in the previous experiment. How many words do you think they encoded on average?

Continue

[Announced Fine]

Part 2 instructions



Task II

In this task you will again encode words using the following Table:

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
8	12	14	10	9	6	24	22	7	5	11	3	18	1	21	16	23	2	13	19	25	4	26	17	20	15

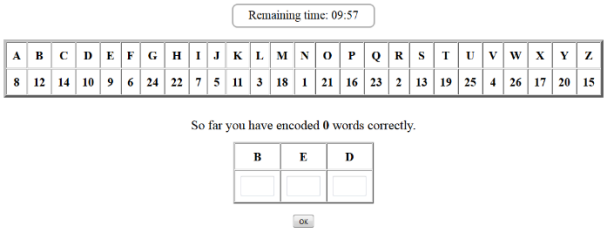
Like in Task I, once you have encoded a word correctly the computer will prompt you with another word to encode. This time, however, you will have **10 minutes** minutes to encode as many words as you want.

For Task II you will receive an initial payment of $2.00. If you encode fewer than 45 words we will reduce your payment by $1.50 (so you will receive $0.50 in total).

When you are ready to start Task II, click continue.

Continue

Part 2 task



Remaining time: 09:57

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
8	12	14	10	9	6	24	22	7	5	11	3	18	1	21	16	23	2	13	19	25	4	26	17	20	15

So far you have encoded **0** words correctly.

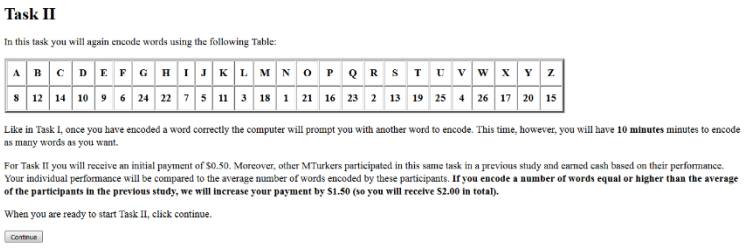
B	E	D

OK

You will receive an initial payment of $2.00. If you encode fewer than 45 words, this payment reduces to $0.50.

[Unannounced Bonus]

Part 2 instructions



Task II

In this task you will again encode words using the following Table:

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
8	12	14	10	9	6	24	22	7	5	11	3	18	1	21	16	23	2	13	19	25	4	26	17	20	15

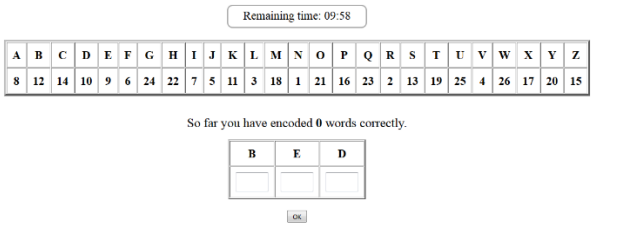
Like in Task I, once you have encoded a word correctly the computer will prompt you with another word to encode. This time, however, you will have **10 minutes** minutes to encode as many words as you want.

For Task II you will receive an initial payment of $0.50. Moreover, other MTurkers participated in this same task in a previous study and earned cash based on their performance. Your individual performance will be compared to the average number of words encoded by these participants. **If you encode a number of words equal or higher than the average of the participants in the previous study, we will increase your payment by $1.50 (so you will receive $2.00 in total).**

When you are ready to start Task II, click continue.

Continue

Part 2 task



Remaining time: 09:58

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
8	12	14	10	9	6	24	22	7	5	11	3	18	1	21	16	23	2	13	19	25	4	26	17	20	15

So far you have encoded **0** words correctly.

B	E	D

OK

You will receive an initial payment of $0.50. If you encode more words than the average of the previous study, this payment increases to $2.00.

Part 2 belief

Before continuing, we ask you to guess the average number of words encoded by the participants in the previous experiment. How many words do you think they encoded on average?

Continue

[Announced Bonus]

Part 2 instructions

Task II

In this task you will again encode words using the following Table:

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
8	12	14	10	9	6	24	22	7	5	11	3	18	1	21	16	23	2	13	19	25	4	26	17	20	15

Like in Task I, once you have encoded a word correctly the computer will prompt you with another word to encode. This time, however, you will have **10 minutes** minutes to encode as many words as you want.

For Task II you will receive an initial payment of $0.50. If you encode 45 words or more, we will increase your payment by $1.50 (so you will receive $2.00 in total).

When you are ready to start Task II, click continue.

Continue

Part 2 task

Remaining time: 09:58

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
8	12	14	10	9	6	24	22	7	5	11	3	18	1	21	16	23	2	13	19	25	4	26	17	20	15

So far you have encoded **0** words correctly.

B	E	D

OK

You will receive an initial payment of $0.50. If you encode 45 words or more, this payment increases to $2.00.

[Common to all treatments]

Part 3

Task III

For Task III you receive an initial flat fee of $1.20. In this task you have to choose for each of the six bets in the table below whether or not you want to take the bet. In each bet there is a 50% chance of winning money and a 50% chance of losing money.

If Task III is selected for payment one bet will be randomly chosen. If you have rejected that bet your bonus payment from this task will be your initial $1.20. If you have accepted that bet, your bonus payment from this task will be your initial $1.20 plus or minus the outcome of that bet.

Please enter your decisions below.

Bet	Accept	Reject
#1. 50% chance of winning $1.00 and 50% chance of losing $0.20	○	○
#2. 50% chance of winning $1.00 and 50% chance of losing $0.40	○	○
#3. 50% chance of winning $1.00 and 50% chance of losing $0.60	○	○
#4. 50% chance of winning $1.00 and 50% chance of losing $0.80	○	○
#5. 50% chance of winning $1.00 and 50% chance of losing $1.00	○	○
#6. 50% chance of winning $1.00 and 50% chance of losing $1.20	○	○

Submit

APPENDIX C: ANALYSIS OF CONTROL TREATMENT

To test whether subjects in our subject pool react at all to short-term monetary incentives, we conducted an additional treatment (**Flat**) with N = 140 participants. This treatment is identical to our other treatments except for the fact that in Part 2 there was no bonus or penalty for (not) meeting a performance target. Instead, subjects received a flat payment of $0.50, regardless of the number of words encoded correctly.

If effort is costly and if subjects are solely motivated by maximizing their own income, then we should observe zero effort in Part 2 of this treatment. If subjects exhibit intrinsic motivation for providing effort (e.g. because they enjoy the task) or if they exhibit some form of other-regarding preferences towards the experimenter (e.g. gift-exchange hypothesis), or other types of reputational considerations (e.g. they worry that if they do not work the experimenter will rate them negatively at the end of the study), then positive effort levels are possible. Yet, we expect these effects to be weaker than the monetary incentives we provided in our Bonus and Penalty treatments. We thus predict that performance, measured as the number of encoded words in Part 2, is lower in our Flat treatment compared to our Bonus and Penalty treatments. We expect no such differences for Part 1 as this was identical for all treatments (subjects always received a piece-rate for words encoded correctly).

These hypotheses are supported by the data. Table C1 compares the work performance for Part 1 and Part 2 across the Flat treatment and our Bonus and Penalty treatments. As we found no differences across Bonus and Penalty, in the following we pool data from these treatments. We also pool across our Announced/Unannounced conditions, but the results hold if we compare Flat separately with each treatment.

As expected, in Part 1 performance is very similar and statistically indistinguishable across treatments. This is further illustrated in Figure C1, showing a substantial overlapping of the cumulative distribution functions (CDFs). This is important because it shows that workers recruited for the Flat and Bonus/Penalty treatments are not different per se in terms of productivity.

In Part 2, in contrast, we find large differences between Flat and the Bonus/Penalty treatments. While in the former participants encode on average around 25 words, in Bonus/Penalty they encode on average 39 words. This difference is highly significant both according to a Mann-Whitney U test as well as a Kolmogorov-Smirnov test. Figure C2 corroborates this finding, showing a stark shift in the cumulative distribution function. In

Flat, it further reveals that while there are some participants who indeed put in very little effort (24% of the participants encoded less than 10 words compared to 2% in the Bonus and Penalty treatment), a majority of subjects put in a positive amount of effort even in the absence of monetary incentives. As explained above, this observation is well in line with previous evidence showing that many people are not solely motivated be pecuniary incentives but also by, e.g., intrinsic motivations or other-regarding preferences.

Table C1 – Performance across the Flat and the Bonus/Penalty treatments

	Bonus / Penalty treatments (n = 853)	**Flat treatment (n = 140)**	**Mann-Whitney U test**	**Kolmogorov-Smirnov test**
Num. words encoded in Part 1	17.45 (5.62)	17.84 (4.77)	$p = 0.289$	$p = 0.795$
Num. words encoded in Part 2	39.33 (11.75)	24.58 (16.33)	$p < 0.001$	$p < 0.001$

Note: The table shows the average number of words encoded correctly. Standard deviation are in parentheses.

Figure C1 – Performance across treatments in Part 1

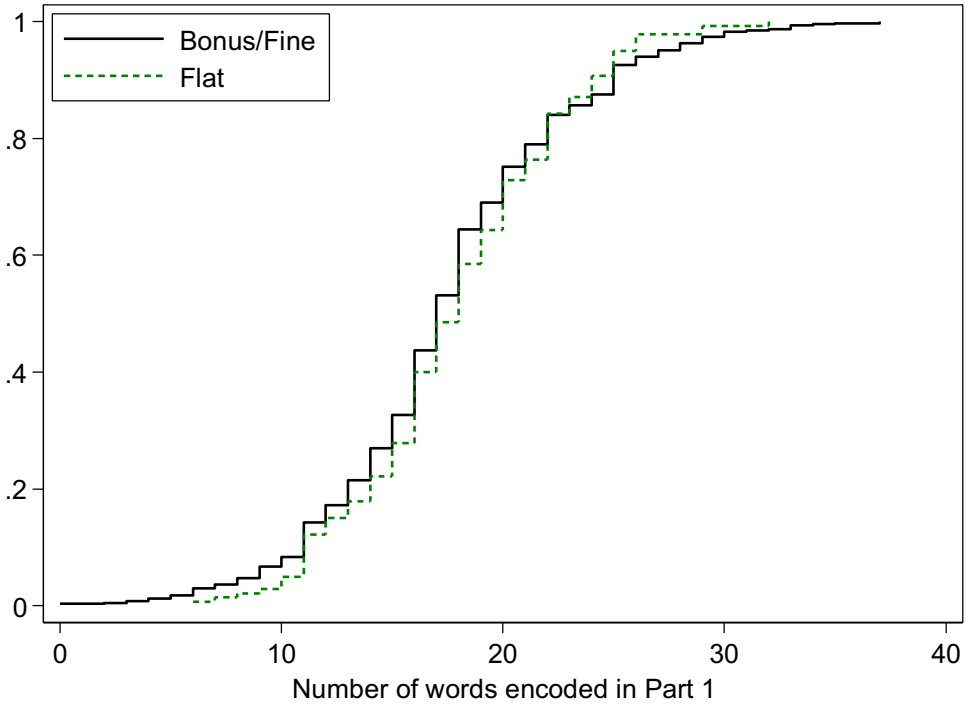
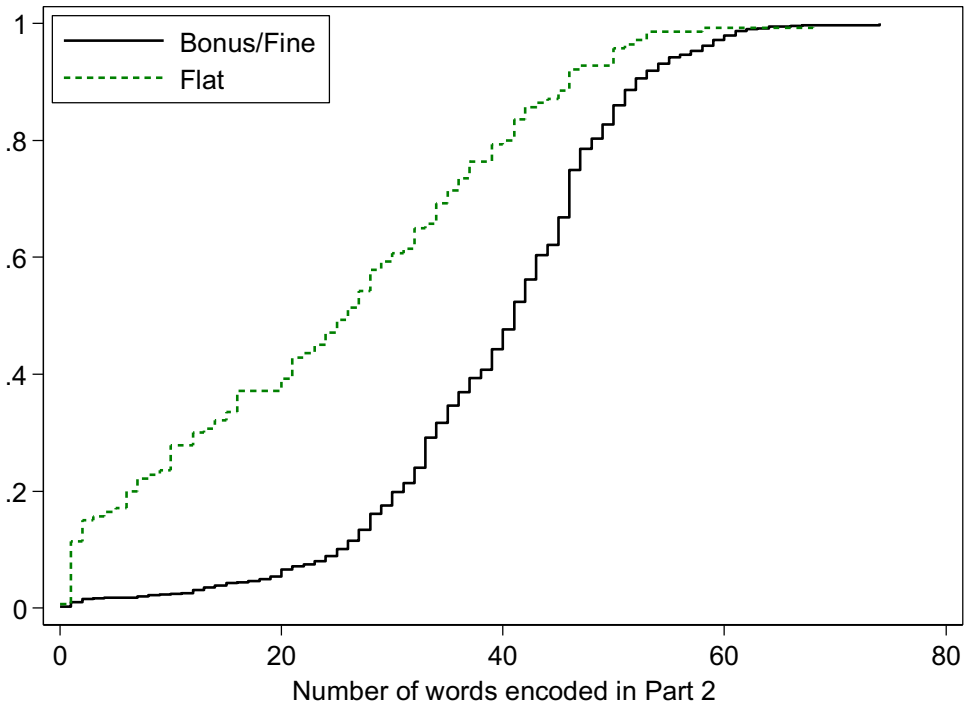


Figure C2 – Performance across treatments in Part 2



APPENDIX D: ADDITIONAL ANALYSES

D.1 Regression analysis of contract framing effects

Table D1 reports an OLS regression of performance in Part 2 of the experiment for the Unannounced and Announced conditions. In both conditions, we regress performance (measured as number of words encoded in Part 2) on a treatment dummy Penalty (equal to 1 for subjects in the Penalty treatment), on a treatment dummy Unannounced (equal to 1 for subjects in the Unannounced treatment), an interaction term between the Penalty and Unannounced dummies, a control of individual ability (measured as number of words encoded in Part 1), a gender dummy, age, and a self-assessment of risk attitudes (measured using the SOEP general risk question).

Table D1 – OLS regressions of performance across treatments

	(1)
Penalty	-0.06
	(0.72)
Unannounced	1.56*
	(0.91)
Penalty x Unannounced	0.13
	(1.28)
Num. Words in Part 1	1.47***
	(0.06)
1 if Female	0.33
	(0.61)
Age	-0.06**
	(0.03)
Risk Loving	-0.05
	(0.11)
Constant	15.15***
	(1.83)
R^2	0.512
N	744

Note: Dependent variable is the number of words encoded in Part 2 of the experiment. For 4 subjects we have missing data on some questionnaire measurements and so they drop out of the regression analysis. Significance levels: *** = 1%; ** = 5%; * = 10%.

The regression confirms that the overall effect of contract framing is small and statistically insignificant in both the Unannounced and Announced conditions. Moreover, the insignificant interaction term indicates that there is no difference between framing effects in the Unannounced and Announced conditions. Among the controls, subjects' ability in the encoding task is strongly and positively associated with performance in Part 2. Furthermore, older subjects tend to encode fewer words.

D.2 Contract framing effects in sub-samples: loss aversion

One common explanation for contract framing effects is loss aversion. In the following analysis we examine whether the effect of framing varies with subjects' degree of loss aversion.

We start by reporting the distribution of loss aversion in our sample. We use the number of lotteries that a subject rejected in Part 3 of the experiment and compute his/her implied degree of loss aversion λ (see Gächter et al. 2010 for details). Table D2 shows that a large fraction of subjects (63%) have $\lambda > 2$, indicating substantial loss aversion in our sample. However, we also find heterogeneity, with 22% of subjects exhibiting a λ close to (or smaller than) 1.

Table D2 – Distribution of loss aversion in the sample

λ	Unannounced	Announced	Overall
> 5	10%	8%	8%
5	22%	28%	26%
2.5	31%	27%	29%
1.67	15%	15%	15%
1.25	8%	11%	10%
1	5%	6%	6%
≤ 0.83	9%	5%	6%

Note: The computation of λ is based on the number of lotteries rejected in Part 3, excluding subjects with multiple switch-points between accepting and rejecting lotteries (9% of subjects). The values of λ are calibrated using the benchmark parameters in Gächter et al. (2010).

Figures D1 and D2 show the cumulative distribution functions (CDFs) of performance in Part 2 of the experiment in the Unannounced and Announced conditions, disaggregated by subject's degree of loss aversion. In particular, we split our sample into two subgroups based on the median degree of loss aversion observed in the data ($\lambda = 2.5$). The left panel of each

Figure shows performance for the above-median subgroups, while the right panel shows the performance of the below-median subgroups.

Figure D1 – Performance in Unannounced for below and above median loss averse subjects in Part 2

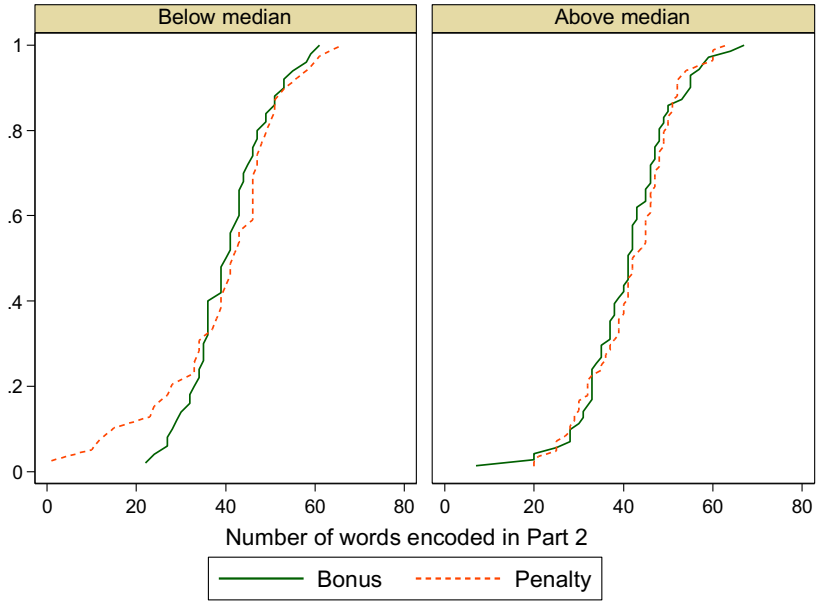
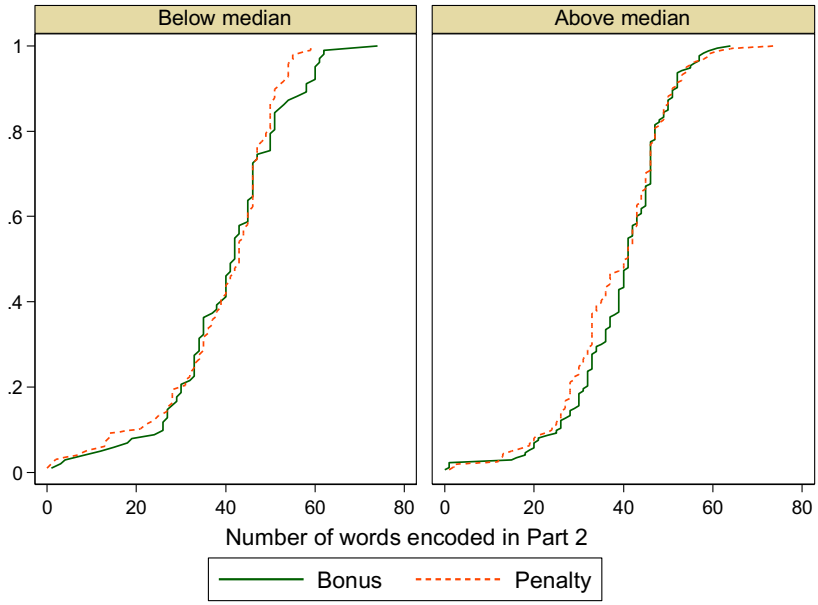


Figure D2 – Performance in Announced for below and above median loss averse subjects in Part 2



For both subgroups, and both in Unannounced and Announced, the CDFs of Bonus and Penalty overlap substantially, indicating very small differences in performance. In

Unannounced, below-median subjects in the Penalty treatment encoded on average 39.62 words (s.d. = 14.01) compared to 40.56 words (s.d. = 9.08) in Bonus. The difference is statistically insignificant ($p = 0.741$ using a two-sided Mann-Whitney test; $p = 0.661$ using a two-sided Kolmogorov-Smirnov test). In Announced, below-median subjects encoded on average 38.96 words (s.d. = 13.00) in the Penalty treatment and 40.27 words (s.d. = 13.46) in Bonus. This difference is also insignificant ($p = 0.853$ using a two-sided Mann-Whitney test; $p = 0.616$ using a two-sided Kolmogorov-Smirnov test).

Above-median subjects in Unannounced encoded on average 41.81 words (s.d. = 9.66) in Penalty and 41.14 words (s.d. = 10.37) in Bonus. The difference is insignificant ($p = 0.556$ using a two-sided Mann-Whitney test; $p = 0.762$ using a two-sided Kolmogorov-Smirnov test). In Announced, above-median subjects encoded on average 38.15 words (s.d. = 12.09) in Penalty and 39.27 words (s.d. = 11.27) in Bonus. The difference is also insignificant ($p = 0.349$ using a two-sided Mann-Whitney test; $p = 0.337$ using a two-sided Kolmogorov-Smirnov test).

Finally, in Table D3 we report OLS regressions of performance in Part 2 (measured as number of words encoded correctly) on a treatment dummy (1 if Penalty treatment), our measure of loss aversion (measured as the number of rejected lotteries in Part 3), and an interaction term between the treatment dummy and the loss aversion measure. We report separate regressions for the Announced and Unannounced treatments. In Unannounced, we also include subjects' beliefs about the target as a regressor. Both regressions include additional controls for individual ability (number of words encoded in Part 1), gender, age, and risk attitudes (measured using the SOEP general risk question).

In both conditions we find that penalties have no impact among the least loss averse subjects (the Penalty dummy is insignificant in both Announced and Unannounced regressions). A subject's degree of loss aversion has no impact on performance under bonus contracts, whereas under penalty contracts loss aversion has a positive effect on effort, and this is weakly significant in Unannounced. The direction of these effects is broadly consistent with the notion that loss aversion may play a role in explaining contract framing effects. However, the frame of the contract seems to have only a limited impact on performance in our setting.

Table D3 – OLS regressions of performance

	Unannounced treatments	Announced treatments
Penalty	-2.86	-2.94
	(1.90)	(2.00)
Loss Aversion	0.03	-0.24
	(0.29)	(0.35)
Penalty x Loss Aversion	0.83*	0.78
	(0.48)	(0.50)
Belief about target	0.22***	
	(0.03)	
Num. Words in Part 1	1.27***	1.47***
	(0.08)	(0.07)
Female	1.40*	-0.12
	(0.78)	(0.81)
Age	-0.06	-0.07*
	(0.04)	(0.04)
Risk Loving	0.21	-0.07
	(0.14)	(0.15)
Constant	11.12***	16.80***
	(2.60)	(2.76)
R^2	0.715	0.464
N	243	531

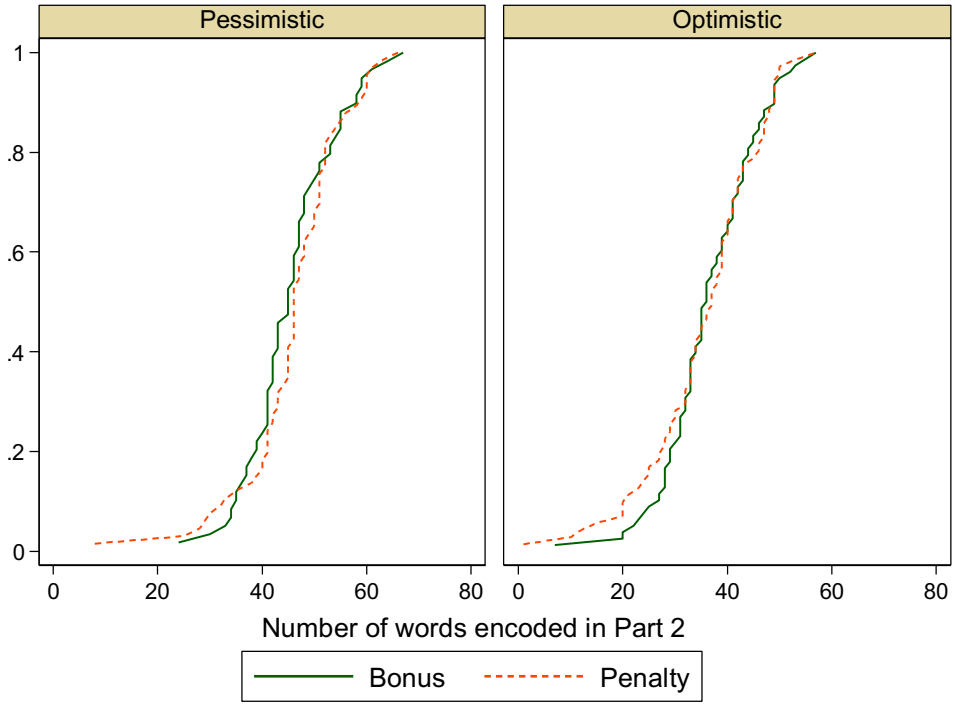
Note: Dependent variable is number of words encoded in Part 2. Loss aversion is measured as the number of rejected lotteries in Part 3. Subjects with multiple switch-points in the lottery task are excluded. For 4 subjects we have missing data on some control variables and so they drop out of the regressions. Standard errors in parentheses. Significance levels: *** = 1%; ** = 5%; * = 10%.

D.3 Contract framing effects in sub-samples: optimism

It is also interesting to check whether the effect of contract framing in the Unannounced condition may depend on subjects' beliefs about the target. Figure D3 shows the cumulative distribution functions (CDFs) of the numbers of performance in Part 2 of the experiment in the Unannounced condition, disaggregated depending on subjects' beliefs about the target. In particular, we split our sample into two subgroups based on the median reported belief of the target (30 words). The left panel of each Figure shows performance for subjects who have

"pessimistic" beliefs, i.e. subjects who believe that the target is relatively high, whereas the panel on the right shows performance for subjects who have relatively "optimistic" beliefs.

Figure D3 – Performance in Unannounced for optimistic and pessimistic subjects with regard to the target in Part 2



Subjects with pessimistic beliefs encoded on average 46.11 words (s.d. = 9.64) in Penalty and 45.49 words (s.d. = 8.53) in Bonus. The difference is insignificant (p = 0.331 using a two-sided Mann-Whitney test; p = 0.580 using a two-sided Kolmogorov-Smirnov test). Subjects with optimistic beliefs encoded on average 35.68 words (s.d. = 10.78) in Penalty and 36.64 words (s.d. = 8.94) in Bonus. The difference is also insignificant (p = 0.873 using a two-sided Mann-Whitney test; p = 0.965 using a two-sided Kolmogorov-Smirnov test).

Table D4 reports regression of number of words encoded in Part 2 among the two subgroups of subjects who hold pessimistic and optimistic beliefs about the target. The regressions include the usual controls. In both subgroups the penalty frame does not have a significant effect on performance.

Table D4 – OLS regressions of performance for optimistic and pessimistic subjects with regard to the target in Unannounced

	Unannounced treatment	
	Pessimistic	Optimistic
Penalty	1.08	-0.98
	(0.89)	(1.12)
Num. Words in Part 1	1.30***	1.45***
	(0.08)	(0.12)
Female	-0.33	1.88
	(0.98)	(1.14)
Age	-0.06	-0.01
	(0.05)	(0.05)
Risk Loving	-0.05	0.25
	(0.16)	(0.20)
Constant	22.19***	11.81***
	(2.93)	(3.59)
R^2	124	147
N	0.713	0.538

Note: Dependent variable is number of words encoded in Part 2. The first model includes only pessimistic subjects who believe that the target is above the median belief of 30. The second model includes optimistic subjects who have below-median beliefs. For a few subjects, we have missing data on some control variables and so they drop out of the regressions. Significance levels: *** = 1%; ** = 5%; * = 10%.