Supplementary Materials

S.1 Meta-analysis

In Figure S.1, we present the forest plot from Figure 1 alongside the data that underlie the information displayed. Both figures were produced using the R package *meta* (Schwarzer et al., 2015). Per Higgins et al. (2021), the SDs for Levitt et al. are adjusted to account for clustered errors, by teacher.

Study	Loss Frame Obs	Loss Frame Mean	Loss Frame SD	Gain Frame Obs	Gain Frame Mean	Gain Frame SD	SMD	95%-CI	Weight
Field									
Pierce et al.	140	0.91	6.60	154	3.23	8.01	-0.313	[-0.543; -0.082]	4.4%
McEvoy	111	81.67	8.05	66	82.19	9.45	-0.060	[-0.365; 0.244]	3.8%
Fryer et al. "Team"	522	0.05	4.84	490	-0.04	5.25	0.017	[-0.106; 0.141]	5.0%
Hossain and List "Team"	47	6.47	0.34	47	6.46	0.34	0.030	[-0.374; 0.435]	3.2%
Hossain and List "Indiv"	53	4.37	0.37	45	4.36	0.36	0.030	[-0.367; 0.428]	3.2%
Levitt et al. "Financial"	948	0.10	0.60	1130	0.08	0.59	0.031	[-0.056; 0.117]	5.2%
Fryer et al. "Indiv"	522	0.16	3.37	483	0.01	4.06	0.040	[-0.083; 0.164]	5.0%
Levitt et al. "Non-Financial"	780	0.06	0.61	667	0.03	0.54	0.050	[-0.054; 0.153]	5.1%
Hong et al.	27	6.32	0.35	27	6.30	0.35	0.066	[-0.468; 0.599]	2.4%
Apostolova-Mihaylova et al.	81	76.30	8.36	90	75.23	8.34	0.128	[-0.173; 0.428]	3.9%
Random effects model	**3231**			**3199**			**0.018**	**[-0.034; 0.070]**	**41.2%**
Heterogeneity: $I^2 = 3\%$ [0%; 64%], $\tau^2 = 0.0004$, $p = 0.41$									
Lab in the Field									
de Quidt et al. "Announced"	287	38.00	12.40	292	39.00	12.00	-0.082	[-0.245; 0.081]	4.8%
Lagarde and Blaauw	60	117.83	34.84	60	116.62	44.28	0.030	[-0.328; 0.388]	3.5%
DellaVigna and Pope	532	2155.00	532.57	545	2136.00	575.69	0.034	[-0.085; 0.154]	5.0%
de Quidt et al. "Unannounced"	137	41.00	11.50	137	40.00	9.78	0.093	[-0.144; 0.330]	4.3%
Brooks et al.	54	14.30	8.10	50	10.40	7.30	0.501	[0.110; 0.892]	3.2%
Armantier and Boly "Unconventional"	34	83.73	6.28	29	80.59	4.82	0.547	[0.042; 1.052]	2.6%
Goldsmith and Dhar	46	8.02	3.49	46	5.65	3.24	0.698	[0.276; 1.119]	3.0%
Random effects model	**1150**			**1159**			**0.207**	**[-0.015; 0.429]**	**26.5%**
Heterogeneity: $I^2 = 71\%$ [36%; 87%], $\tau^2 = 0.0641$, $p < 0.01$									
Laboratory									
Dolan et al.	789	508.62	228.24	784	530.74	219.38	-0.099	[-0.198; 0.000]	5.1%
Grolleau et al.	150	3.45	2.27	150	3.61	2.48	-0.067	[-0.294; 0.159]	4.4%
Church et al.	31	1.14	0.17	36	1.11	0.34	0.108	[-0.373; 0.588]	2.7%
Brooks et al. (Stated)	73	1.86	1.16	72	1.58	0.97	0.260	[-0.067; 0.587]	3.7%
Armantier and Boly "Conventional"	56	82.31	7.80	58	78.68	7.88	0.460	[0.088; 0.832]	3.4%
Imas et al.	43	17.28	3.33	40	15.35	4.50	0.486	[0.049; 0.923]	3.0%
Hannan et al. (Stated)	33	9.58	3.31	35	7.40	4.58	0.537	[0.052; 1.021]	2.7%
Imas et al. "Pilot"	30	15.27	4.44	32	11.88	5.55	0.664	[0.151; 1.176]	2.5%
Bulte et al. "Task1, envelope folding"	800	27.63	8.78	200	20.57	6.71	0.840	[0.680; 0.999]	4.8%
Random effects model	**2005**			**1407**			**0.334**	**[0.108; 0.560]**	**32.3%**
Heterogeneity: $I^2 = 93\%$ [88%; 95%], $\tau^2 = 0.0882$, $p < 0.01$									
Random effects model	**6386**			**5765**			**0.160**	**[0.048; 0.272]**	**100.0%**
Heterogeneity: $I^2 = 84\%$ [77%; 88%], $\tau^2 = 0.0611$, $p < 0.01$									
Test for subgroup differences: $\chi^2_2 = 9.38$, df = 2 ($p < 0.01$)									

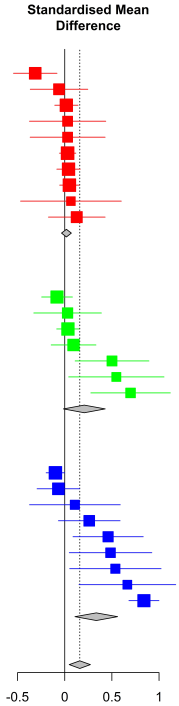

Fig. S.1 Meta-analysis of experimental studies estimating the effect of loss-framed contracts on productivity (effort). For each study, the figure reports the number of observations (Obs), the mean value of the outcome measure in units reported in the experiment (Mean), and the standard deviation of the outcome measure (SD) for the loss-framed-contract and gain-framed-contract groups. It also reports the standardized mean difference (SMD) between the outcomes in the loss-framed-contract and gain-framed-contract groups (standardized by the pooled standard deviation) and its 95% CI. The final column reports the weight that each study contributes to the summary estimated treatment effect.

Table S.1 Excluded estimates within included publications

Study	Excluded	Reason
Goldsmith and Dhar (2011)	Exp 1A	The task was nearly impossible and designed to increase time spent on task w/o increasing productivity.
Goldsmith and Dhar (2011)	Exps 2-4	Surveys about how frames are perceived.
Armantier and Boly (2012)	Time	Not a productivity measure.
Imas et al. (2016)	Exp 2	Subjects selected into participation through WTP.
Brooks et al. (2017)	Low-Bar & Extreme	Were designed to demonstrate prepaying for too few (too many) units can harm productivity.
Bulte et al. (2020)	Task 2	Subjects selected desired frame.

S.1.1 Tests of asymmetry in funnel plots

The Begg-Mazumdar method tests if the rank according to effect size is correlated with the rank according to standard error size (Begg and Mazumdar, 1994). The test statistic, z, is the difference in pairs with positive correlation and the pairs of with negative correlation, adjusted for number of studies. In expectation, it is mean zero. The probability of the calculated value in a standard normal distribution is the test's $p - value$. For the studies in the meta-analysis $z = 2.47, p - value = 0.01$.

The Thompson-Sharp test regresses each study's z-scored SMD on the inverse of its standard error (Thompson and Sharp, 1999). It assumes that studies with fewer observations and smaller inverse standard errors should not systematically have larger standardized effects, and thus the regression estimate of the constant term should be zero. The test statistic of the constant term is calculated as it normally is in a regression–via a t-distribution. The estimated constant for our studies is $t = 4.33, p - value = 0.0007$.

The trim-and-fill technique uses an iterative method that first trims the studies that lie outside the expected error range. Then the “true” center of the funnel is then estimated from the remaining subset. Then, for each trimmed study, a “fill” study is added to the full set of studies. The filled study is the same distance from the true center as the trimmed study, but on the opposite side of the funnel. The final step is to asses if symmetry has been achieved; if it has not, there is another iteration. By default, meta uses the “L” method (Duval and Tweedie, 2000a,b), which employs the fixed-effects model estimator of the true effect, and uses a rank statistic to trim the asymmetric studies. It then uses estimate random-effects model to assess if symmetry has been achieved.

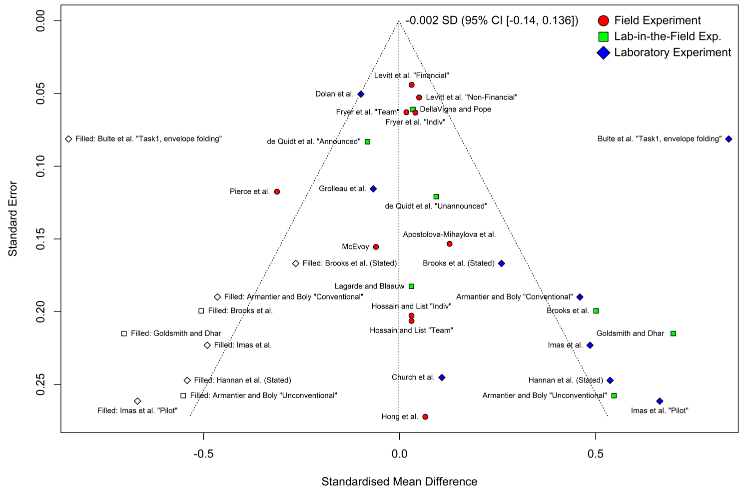

Fig. S.2 Funnel plot after studies are trimmed and filled. Estimated standardized effect sizes and their standard errors (black shapes) plus counterfactual studies (white shapes) that are added by a "trim-and-fill" approach to generate a more symmetric funnel. The dotted vertical line is the revised summary estimated effect from loss-framed contracts.

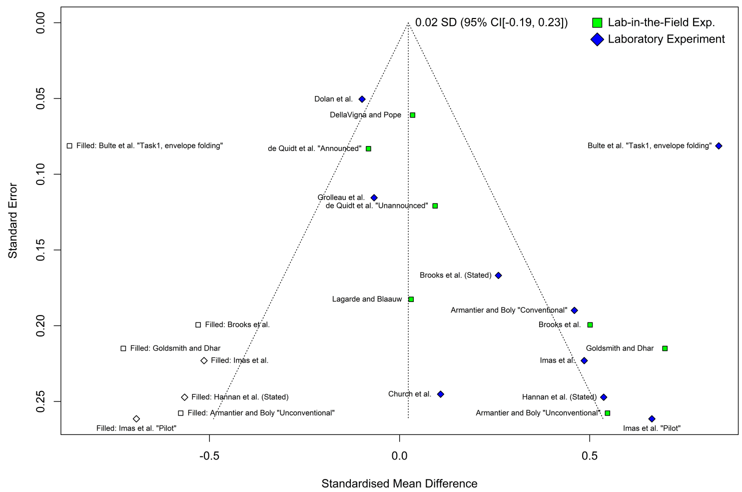

Fig. S.3 Funnel plot after laboratory studies are trimmed and filled. Estimated standardized effect sizes and their standard errors (black shapes) plus counterfactual studies (white shapes) that are added by a "trim-and-fill" approach to generate a more symmetric funnel. The dotted vertical line is the revised summary estimated effect from loss-framed contracts.

S.1.2 Piece-rate versus threshold

Figure S.4 presents a forest plot of the same studies as Figure S.1, except subgroups are divided by contract type (piece-rate or threshold), rather than being divided by setting as in Figure S.1. For piece-rate contracts, the summary estimated effect size is 0.15 SD (95% CI [-0.01, 0.31]). For threshold contracts, the summary estimated effect size is: 0.17 SD (95% CI [0.01, 0.33]).

Designing the right threshold contract to motivate workers is difficult. To illustrate this difficulty, we use the example from the first paragraph above, where the worker was producing 15 units, and suppose she can increase her output a maximum of 3 more units. She would not be motivated by the threshold contract, but she would be motivated by the piece-rate contract; even producing 18 units, she would receive the same $10 penalty under the threshold contract, but her penalty would be reduced from $5 to $2, under piece-rate. Thus, she would choose not to produce the additional 3 units under threshold, but would under piece-rate.

To motivate this worker, the threshold could be set to 18, rather than 20. Then the threshold contract would more strongly motivate her than the piece-rate contract ($10 is more than $3). However, if worker output is heterogeneous, a threshold that motivated this worker might be too high or low for her peers. A coworker whose natural output is only 12 might only be able to increase his output to 14, so will not be motivated by that threshold. Another coworker, who naturally produces 18, might have increased output to 20 under piece-rate or had the threshold been 20, might only produce 18, if the threshold were reduced. Brooks et al. (2017) showed that even under piece-rate the quota may be too high or low to be effective, but piece-rate seems more robust to heterogeneous abilities and misestimation of baseline productivity.

Study	Loss Frame Obs	Loss Frame Mean	Loss Frame SD	Gain Frame Obs	Gain Frame Mean	Gain Frame SD	SMD	95%-CI	Weight
Piece-rate									
Dolan et al.	789	508.62	228.24	784	530.74	219.38	-0.099	[-0.198; 0.000]	5.1%
Grolleau et al.	150	3.45	2.27	150	3.61	2.48	-0.067	[-0.294; 0.159]	4.4%
McEvoy	111	81.67	8.05	66	82.19	9.45	-0.060	[-0.365; 0.244]	3.8%
Fryer et al. "Team"	522	0.05	4.84	490	-0.04	5.25	0.017	[-0.106; 0.141]	5.0%
Lagarde and Blaauw	60	117.83	34.84	60	116.62	44.28	0.030	[-0.328; 0.388]	3.5%
Fryer et al. "Indiv"	522	0.16	3.37	483	0.01	4.06	0.040	[-0.083; 0.164]	5.0%
Apostolova-Mihaylova et al.	81	76.30	8.36	90	75.23	8.34	0.128	[-0.173; 0.428]	3.9%
Armantier and Boly "Conventional"	56	82.31	7.80	58	78.68	7.88	0.460	[0.088; 0.832]	3.4%
Brooks et al.	54	14.30	8.10	50	10.40	7.30	0.501	[0.110; 0.892]	3.2%
Armantier and Boly "Unconventional"	34	83.73	6.28	29	80.59	4.82	0.547	[0.042; 1.052]	2.6%
Goldsmith and Dhar	46	8.02	3.49	46	5.65	3.24	0.698	[0.276; 1.119]	3.0%
Random effects model	**2425**			**2306**			**0.147**	**[-0.013; 0.308]**	**43.0%**
Heterogeneity: $I^2 = 69\%$ [42%; 83%], $\tau^2 = 0.0517$, $p < 0.01$									
Threshold									
Pierce et al.	140	0.91	6.60	154	3.23	8.01	-0.313	[-0.543; -0.082]	4.4%
de Quidt et al. "Announced"	287	38.00	12.40	292	39.00	12.00	-0.082	[-0.245; 0.081]	4.8%
Hossain and List "Team"	47	6.47	0.34	47	6.46	0.34	0.030	[-0.374; 0.435]	3.2%
Hossain and List "Indiv"	53	4.37	0.37	45	4.36	0.36	0.030	[-0.367; 0.428]	3.2%
Levitt et al. "Financial"	948	0.10	0.60	1130	0.08	0.59	0.031	[-0.056; 0.117]	5.2%
DellaVigna and Pope	532	2155.00	532.57	545	2136.00	575.69	0.034	[-0.085; 0.154]	5.0%
Levitt et al. "Non-Financial"	780	0.06	0.61	667	0.03	0.54	0.050	[-0.054; 0.153]	5.1%
Hong et al.	27	6.32	0.35	27	6.30	0.35	0.066	[-0.468; 0.599]	2.4%
de Quidt et al. "Unannounced"	137	41.00	11.50	137	40.00	9.78	0.093	[-0.144; 0.330]	4.3%
Church et al.	31	1.14	0.17	36	1.11	0.34	0.108	[-0.373; 0.588]	2.7%
Brooks et al. (Stated)	73	1.86	1.16	72	1.58	0.97	0.260	[-0.067; 0.587]	3.7%
Imas et al.	43	17.28	3.33	40	15.35	4.50	0.486	[0.048; 0.923]	3.0%
Hannan et al. (Stated)	33	9.58	3.31	35	7.40	4.58	0.537	[0.052; 1.021]	2.7%
Imas et al. "Pilot"	30	15.27	4.44	32	11.88	5.55	0.664	[0.151; 1.176]	2.5%
Bulte et al. "Task1, envelope folding"	800	27.63	8.78	200	20.57	6.71	0.840	[0.680; 0.999]	4.8%
Random effects model	**3961**			**3459**			**0.168**	**[0.009; 0.327]**	**57.0%**
Heterogeneity: $I^2 = 88\%$ [82%; 92%], $\tau^2 = 0.0725$, $p < 0.01$									
Random effects model	**6386**			**5765**			**0.160**	**[0.048; 0.272]**	**100.0%**
Heterogeneity: $I^2 = 84\%$ [77%; 88%], $\tau^2 = 0.0611$, $p < 0.01$									
Test for subgroup differences: $\chi^2_1 = 0.03$, df = 1 ($p = 0.86$)									

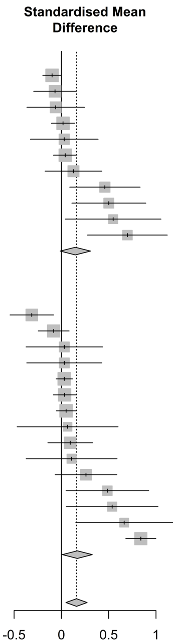

Fig. S.4 Meta-analysis of experimental studies of loss-framed contracts grouped by piece-rate versus threshold designs

S.1.3 Advance payment

Figure S.5 presents a forest, plot which tests for difference between experiments in which the workers received the reward (payment) in advance versus studies in which they were merely told they would get the reward. While the estimated effect size is larger when the workers get the reward in advance 0.24 SD (95% CI [0.01, 0.46]) than when the do not 0.08 SD (95% CI [-0.02, 0.17]), the difference is not statistically significant ($\chi^2 = 1.63$, $df = 1$, $p = 0.20$).

Study	Loss Frame Obs	Loss Frame Mean	Loss Frame SD	Gain Frame Obs	Gain Frame Mean	Gain Frame SD	SMD	95%-CI	Weight
NO Advance Payment									
Dolan et al.	789	508.62	228.24	784	530.74	219.38	-0.099	[-0.198; 0.000]	5.1%
de Quidt et al. "Announced"	287	38.00	12.40	292	39.00	12.00	-0.082	[-0.245; 0.081]	4.8%
McEvoy	111	81.67	8.05	66	82.19	9.45	-0.060	[-0.365; 0.244]	3.8%
Lagarde and Blaauw	60	117.83	34.84	60	116.62	44.28	0.030	[-0.328; 0.388]	3.5%
Hossain and List "Team"	47	6.47	0.34	47	6.46	0.34	0.030	[-0.374; 0.435]	3.2%
Hossain and List "Indiv"	53	4.37	0.37	45	4.36	0.36	0.030	[-0.367; 0.428]	3.2%
DellaVigna and Pope	532	2155.00	532.57	545	2136.00	575.69	0.034	[-0.085; 0.154]	5.0%
Hong et al.	27	6.32	0.35	27	6.30	0.35	0.066	[-0.468; 0.599]	2.4%
de Quidt et al. "Unannounced"	137	41.00	11.50	137	40.00	9.78	0.093	[-0.144; 0.330]	4.3%
Church et al.	31	1.14	0.17	36	1.11	0.34	0.108	[-0.373; 0.588]	2.7%
Apostolova-Mihaylova et al.	81	76.30	8.36	90	75.23	8.34	0.128	[-0.173; 0.428]	3.9%
Brooks et al. (Stated)	73	1.86	1.16	72	1.58	0.97	0.260	[-0.067; 0.587]	3.7%
Armantier and Boly "Conventional"	56	82.31	7.80	58	78.68	7.88	0.460	[0.088; 0.832]	3.4%
Brooks et al.	54	14.30	8.10	50	10.40	7.30	0.501	[0.110; 0.892]	3.2%
Armantier and Boly "Unconventional"	34	83.73	6.28	29	80.59	4.82	0.547	[0.042; 1.052]	2.6%
Random effects model	**2372**			**2338**			**0.078**	**[-0.017; 0.174]**	**54.8%**
Heterogeneity: $I^2 = 46\%$ [1%; 70%], $\tau^2 = 0.0132$, $p = 0.03$									
Advance Payment									
Pierce et al.	140	0.91	6.60	154	3.23	8.01	-0.313	[-0.543; -0.082]	4.4%
Grolleau et al.	150	3.45	2.27	150	3.61	2.48	-0.067	[-0.294; 0.159]	4.4%
Fryer et al. "Team"	522	0.05	4.84	490	-0.04	5.25	0.017	[-0.106; 0.141]	5.0%
Levitt et al. "Financial"	948	0.10	0.60	1130	0.08	0.59	0.031	[-0.056; 0.117]	5.2%
Fryer et al. "Indiv"	522	0.16	3.37	483	0.01	4.06	0.040	[-0.083; 0.164]	5.0%
Levitt et al. "Non-Financial"	780	0.06	0.61	667	0.03	0.54	0.050	[-0.054; 0.153]	5.1%
Imas et al.	43	17.28	3.33	40	15.35	4.50	0.486	[0.048; 0.923]	3.0%
Hannan et al. (Stated)	33	9.58	3.31	35	7.40	4.58	0.537	[0.052; 1.021]	2.7%
Imas et al. "Pilot"	30	15.27	4.44	32	11.88	5.55	0.664	[0.151; 1.176]	2.5%
Goldsmith and Dhar	46	8.02	3.49	46	5.65	3.24	0.698	[0.276; 1.119]	3.0%
Bulte et al. "Task1, envelope folding"	800	27.63	8.78	200	20.57	6.71	0.840	[0.680; 0.999]	4.8%
Random effects model	**4014**			**3427**			**0.237**	**[0.012; 0.462]**	**45.2%**
Heterogeneity: $I^2 = 92\%$ [87%; 95%], $\tau^2 = 0.1236$, $p < 0.01$									
Random effects model	**6386**			**5765**			**0.160**	**[0.048; 0.272]**	**100.0%**
Heterogeneity: $I^2 = 84\%$ [77%; 88%], $\tau^2 = 0.0611$, $p < 0.01$									
Test for subgroup differences: $\chi^2_1 = 1.63$, df = 1 ($p = 0.20$)									

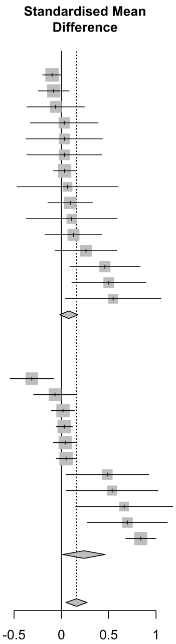

Fig. S.5 Meta-analysis of experimental studies of loss-framed contracts grouped by whether subject received payment in advance

S.1.4 Limiting to studies focusing on loss framing effect on effort

To estimate the effect size of the literature scholars typical associate with loss-framed contracts, we remove two studies from our dataset and present the results of that meta-analysis in Figure S.6. We remove Dolan et al. (2012), because their experiment was in a government report but loss-framing was not the focus of the report. We remove Grolleau et al. (2016) because the central claim of the paper is that loss-framed contracts make cheating more likely, not that the authors fail to detect an effect of loss-framed contracts on effort.

S.1.5 Re-classifying "lab-in-the-field" experiments as "field experiments"

Because some scholars classify laboratory experiments conducted with non-standard subjects as "artefactual" field experiments, we re-do the meta-analysis after re-classifying these lab-in-the-field experiments as "field experiments" and present the results in Figure S.7. Because these experiments have some of the largest effect sizes, the gap between the summary effect sizes for laboratory and field experiment decreases substantially, but field experiments still

Study	Loss Frame Obs	Loss Frame Mean	Loss Frame SD	Gain Frame Obs	Gain Frame Mean	Gain Frame SD	SMD	95%-CI	Weight
Field									
Pierce et al.	140	0.91	6.60	154	3.23	8.01	-0.313	[-0.543; -0.082]	4.8%
McEvoy	111	81.67	8.05	66	82.19	9.45	-0.060	[-0.365; 0.244]	4.2%
Fryer et al. "Team"	522	0.05	4.84	490	-0.04	5.25	0.017	[-0.106; 0.141]	5.6%
Hossain and List "Team"	47	6.47	0.34	47	6.46	0.34	0.030	[-0.374; 0.435]	3.5%
Hossain and List "Indiv"	53	4.37	0.37	45	4.36	0.36	0.030	[-0.367; 0.428]	3.5%
Levitt et al. "Financial"	948	0.10	0.60	1130	0.08	0.59	0.031	[-0.056; 0.117]	5.7%
Fryer et al. "Indiv"	522	0.16	3.37	483	0.01	4.06	0.040	[-0.083; 0.164]	5.6%
Levitt et al. "Non-Financial"	780	0.06	0.61	667	0.03	0.54	0.050	[-0.054; 0.153]	5.7%
Hong et al.	27	6.32	0.35	27	6.30	0.35	0.066	[-0.468; 0.599]	2.7%
Apostolova-Mihaylova et al.	81	76.30	8.36	90	75.23	8.34	0.128	[-0.173; 0.428]	4.3%
Random effects model	**3231**			**3199**			**0.018**	**[-0.034; 0.070]**	**45.6%**
Heterogeneity: $I^2 = 3\%$ [0%; 64%], $\tau^2 = 0.0004$, $p = 0.41$									
Lab in the Field									
de Quidt et al. "Announced"	287	38.00	12.40	292	39.00	12.00	-0.082	[-0.245; 0.081]	5.3%
Lagarde and Blaauw	60	117.83	34.84	60	116.62	44.28	0.030	[-0.328; 0.388]	3.8%
DellaVigna and Pope	532	2155.00	532.57	545	2136.00	575.69	0.034	[-0.085; 0.154]	5.6%
de Quidt et al. "Unannounced"	137	41.00	11.50	137	40.00	9.78	0.093	[-0.144; 0.330]	4.8%
Brooks et al.	54	14.30	8.10	50	10.40	7.30	0.501	[0.110; 0.892]	3.6%
Armantier and Boly "Unconventional"	34	83.73	6.28	29	80.59	4.82	0.547	[0.042; 1.052]	2.8%
Goldsmith and Dhar	46	8.02	3.49	46	5.65	3.24	0.698	[0.276; 1.119]	3.4%
Random effects model	**1150**			**1159**			**0.207**	**[-0.015; 0.429]**	**29.3%**
Heterogeneity: $I^2 = 71\%$ [36%; 87%], $\tau^2 = 0.0641$, $p < 0.01$									
Laboratory									
Church et al.	31	1.14	0.17	36	1.11	0.34	0.108	[-0.373; 0.588]	3.0%
Brooks et al. (Stated)	73	1.86	1.16	72	1.58	0.97	0.260	[-0.067; 0.587]	4.1%
Armantier and Boly "Conventional"	56	82.31	7.80	58	78.68	7.88	0.460	[0.088; 0.832]	3.7%
Imas et al.	43	17.28	3.33	40	15.35	4.50	0.486	[0.049; 0.923]	3.3%
Hannan et al. (Stated)	33	9.58	3.31	35	7.40	4.58	0.537	[0.052; 1.021]	3.0%
Imas et al. "Pilot"	30	15.27	4.44	32	11.88	5.55	0.664	[0.151; 1.176]	2.8%
Bulte et al. "Task1, envelope folding"	800	27.63	8.78	200	20.57	6.71	0.840	[0.680; 0.999]	5.3%
Random effects model	**1066**			**473**			**0.519**	**[0.325; 0.712]**	**25.1%**
Heterogeneity: $I^2 = 66\%$ [23%; 85%], $\tau^2 = 0.0311$, $p < 0.01$									
Random effects model	**5447**			**4831**			**0.186**	**[0.068; 0.304]**	**100.0%**
Test for subgroup differences: $\chi^2_2 = 25.79$, df = 2 ($p < 0.01$)									

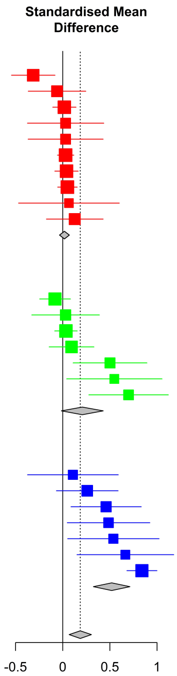

Fig. S.6 Meta-analysis of studies focusing on loss framing effect on effort

have a 95% CI that includes zero (and now the study estimates from field and laboratory experiments are equally heterogeneous).

Study	Loss Frame Obs	Loss Frame Mean	Loss Frame SD	Gain Frame Obs	Gain Frame Mean	Gain Frame SD	SMD	95%-CI	Weight
Field									
Pierce et al.	140	0.91	6.60	154	3.23	8.01	-0.313	[-0.543; -0.082]	4.4%
de Quidt et al. "Announced"	287	38.00	12.40	292	39.00	12.00	-0.082	[-0.245; 0.081]	4.8%
McEvoy	111	81.67	8.05	66	82.19	9.45	-0.060	[-0.365; 0.244]	3.8%
Fryer et al. "Team"	522	0.05	4.84	490	-0.04	5.25	0.017	[-0.106; 0.141]	5.0%
Lagarde and Blaauw	60	117.83	34.84	60	116.62	44.28	0.030	[-0.328; 0.388]	3.5%
Hossain and List "Team"	47	6.47	0.34	47	6.46	0.34	0.030	[-0.374; 0.435]	3.2%
Hossain and List "Indiv"	53	4.37	0.37	45	4.36	0.36	0.030	[-0.367; 0.428]	3.2%
Levitt et al. "Financial"	948	0.10	0.60	1130	0.08	0.59	0.031	[-0.056; 0.117]	5.2%
DellaVigna and Pope	532	2155.00	532.57	545	2136.00	575.69	0.034	[-0.085; 0.154]	5.0%
Fryer et al. "Indiv"	522	0.16	3.37	483	0.01	4.06	0.040	[-0.083; 0.164]	5.0%
Levitt et al. "Non-Financial"	780	0.06	0.61	667	0.03	0.54	0.050	[-0.054; 0.153]	5.1%
Hong et al.	27	6.32	0.35	27	6.30	0.35	0.066	[-0.468; 0.599]	2.4%
de Quidt et al. "Unannounced"	137	41.00	11.50	137	40.00	9.78	0.093	[-0.144; 0.330]	4.3%
Apostolova-Mihaylova et al.	81	76.30	8.36	90	75.23	8.34	0.128	[-0.173; 0.428]	3.9%
Brooks et al.	54	14.30	8.10	50	10.40	7.30	0.501	[0.110; 0.892]	3.2%
Armantier and Boly "Unconventional"	34	83.73	6.28	29	80.59	4.82	0.547	[0.042; 1.052]	2.6%
Goldsmith and Dhar	46	8.02	3.49	46	5.65	3.24	0.698	[0.276; 1.119]	3.0%
Random effects model	**4381**			**4358**			**0.063**	**[-0.036; 0.162]**	**67.7%**
Heterogeneity: $I^2 = 48\%$ [9%; 70%], $\tau^2 = 0.0254$, $p = 0.01$									
Lab									
Dolan et al.	789	508.62	228.24	784	530.74	219.38	-0.099	[-0.198; 0.000]	5.1%
Grolleau et al.	150	3.45	2.27	150	3.61	2.48	-0.067	[-0.294; 0.159]	4.4%
Church et al.	31	1.14	0.17	36	1.11	0.34	0.108	[-0.373; 0.588]	2.7%
Brooks et al. (Stated)	73	1.86	1.16	72	1.58	0.97	0.260	[-0.067; 0.587]	3.7%
Armantier and Boly "Conventional"	56	82.31	7.80	58	78.68	7.88	0.460	[0.088; 0.832]	3.4%
Imas et al.	43	17.28	3.33	40	15.35	4.50	0.486	[0.049; 0.923]	3.0%
Hannan et al. (Stated)	33	9.58	3.31	35	7.40	4.58	0.537	[0.052; 1.021]	2.7%
Imas et al. "Pilot"	30	15.27	4.44	32	11.88	5.55	0.664	[0.151; 1.176]	2.5%
Bulte et al. "Task1, envelope folding"	800	27.63	8.78	200	20.57	6.71	0.840	[0.680; 0.999]	4.8%
Random effects model	**2005**			**1407**			**0.334**	**[0.108; 0.560]**	**32.3%**
Heterogeneity: $I^2 = 93\%$ [88%; 95%], $\tau^2 = 0.0882$, $p < 0.01$									
Random effects model	**6386**			**5765**			**0.160**	**[0.048; 0.272]**	**100.0%**
Test for subgroup differences: $\chi^2_1 = 4.65$, df = 1 ($p = 0.03$)									

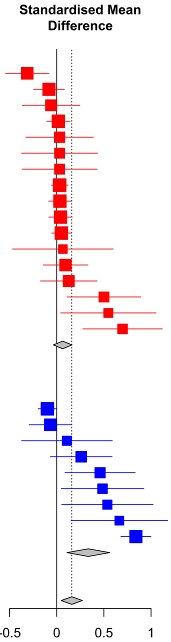

Fig. S.7 Meta-analysis re-classifying "Lab-in-the-field" as "field"

Study	Loss Frame Obs	Loss Frame Mean	Loss Frame SD	Gain Frame Obs	Gain Frame Mean	Gain Frame SD	SMD	95%-CI	Weight
Field									
Pierce et al.	140	0.91	6.60	154	3.23	8.01	-0.313	[-0.543; -0.082]	4.4%
McEvoy	111	81.67	8.05	66	82.19	9.45	-0.060	[-0.365; 0.244]	3.8%
Fryer et al. "Team"	522	0.05	4.84	490	-0.04	5.25	0.017	[-0.106; 0.141]	5.0%
Hossain and List "Team"	47	6.47	0.34	47	6.46	0.34	0.030	[-0.374; 0.435]	3.2%
Hossain and List "Indiv"	53	4.37	0.37	45	4.36	0.36	0.030	[-0.367; 0.428]	3.2%
Levitt et al. "Financial"	948	0.10	0.60	1130	0.08	0.59	0.031	[-0.056; 0.117]	5.2%
Fryer et al. "Indiv"	522	0.16	3.37	483	0.01	4.06	0.040	[-0.083; 0.164]	5.0%
Levitt et al. "Non-Financial"	780	0.06	0.61	667	0.03	0.54	0.050	[-0.054; 0.153]	5.1%
Hong et al.	27	6.32	0.35	27	6.30	0.35	0.066	[-0.468; 0.599]	2.4%
Apostolova-Mihaylova et al.	81	76.30	8.36	90	75.23	8.34	0.128	[-0.173; 0.428]	3.9%
Random effects model	**3231**			**3199**			**0.018**	**[-0.034; 0.070]**	**41.2%**
Heterogeneity: $I^2 = 3\%$ [0%; 64%], $\tau^2 = 0.0004$, $p = 0.41$									
Lab									
Dolan et al.	789	508.62	228.24	784	530.74	219.38	-0.099	[-0.198; 0.000]	5.1%
de Quidt et al. "Announced"	287	38.00	12.40	292	39.00	12.00	-0.082	[-0.245; 0.081]	4.8%
Grolleau et al.	150	3.45	2.27	150	3.61	2.48	-0.067	[-0.294; 0.159]	4.4%
Lagarde and Blaauw	60	117.83	34.84	60	116.62	44.28	0.030	[-0.328; 0.388]	3.5%
DellaVigna and Pope	532	2155.00	532.57	545	2136.00	575.69	0.034	[-0.085; 0.154]	5.0%
de Quidt et al. "Unannounced"	137	41.00	11.50	137	40.00	9.78	0.093	[-0.144; 0.330]	4.3%
Church et al.	31	1.14	0.17	36	1.11	0.34	0.108	[-0.373; 0.588]	2.7%
Brooks et al. (Stated)	73	1.86	1.16	72	1.58	0.97	0.260	[-0.067; 0.587]	3.7%
Armantier and Boly "Conventional"	56	82.31	7.80	58	78.68	7.88	0.460	[0.088; 0.832]	3.4%
Imas et al.	43	17.28	3.33	40	15.35	4.50	0.486	[0.049; 0.923]	3.0%
Brooks et al.	54	14.30	8.10	50	10.40	7.30	0.501	[0.110; 0.892]	3.2%
Hannan et al. (Stated)	33	9.58	3.31	35	7.40	4.58	0.537	[0.052; 1.021]	2.7%
Armantier and Boly "Unconventional"	34	83.73	6.28	29	80.59	4.82	0.547	[0.042; 1.052]	2.6%
Imas et al. "Pilot"	30	15.27	4.44	32	11.88	5.55	0.664	[0.151; 1.176]	2.5%
Goldsmith and Dhar	46	8.02	3.49	46	5.65	3.24	0.698	[0.276; 1.119]	3.0%
Bulte et al. "Task1, envelope folding"	800	27.63	8.78	200	20.57	6.71	0.840	[0.680; 0.999]	4.8%
Random effects model	**3155**			**2566**			**0.279**	**[0.120; 0.437]**	**58.8%**
Heterogeneity: $I^2 = 89\%$ [84%; 92%], $\tau^2 = 0.0757$, $p < 0.01$									
Random effects model	**6386**			**5765**			**0.160**	**[0.048; 0.272]**	**100.0%**
Test for subgroup differences: $\chi^2_1 = 9.39$, df = 1 ($p < 0.01$)									

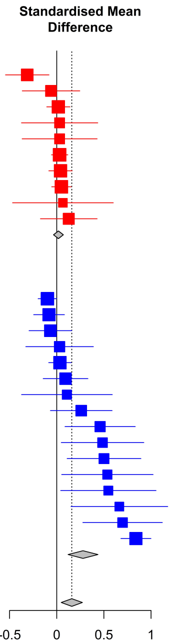

Fig. S.8 Meta-analysis re-classifying "Lab-in-the-field" as "lab"

S.2 Experiment

S.2.1 Preference versus WTP

Here, we justify with a numerical example the claim that "If the margin by which loss-frame-preferring people are willing to pay more for their preferred contract is larger than the margin by which gain-frame-preferring people are willing to pay more for their preferred contract, then it is possible for $Mean(WTP_{LF}) > Mean(WTP_{GF})$ even if most people prefer the gain-framed contract." For example, say that: (1) 60% prefer gain-framed contracts to loss-framed contracts and, for each person, $\$0.85 = WTP_{LF} < WTP_{GF} = \$0.90$; (2) 25% prefer loss-framed contract to gain-framed contracts and, for each person, $\$2 = WTP_{LF} > WTP_{GF} = \$0.50$; and (3) 15% of the people are indifferent and, for each person, $\$1 = WTP_{LF} = WTP_{GF}$. Then, despite only a minority preferring the loss frame, $Mean(WTP_{LF}) > Mean(WTP_{GF})$ i.e., $Mean(WTP_{LF}) = .15 * 1 + .6 * 0.85 + .25 * 2 = \$1.16 > Mean(WTP_{GF}) = .15 * 1 + .6 * .90 + .25 * .5 = \$0.815$.

S.2.2 Alternative regression specifications

Tables S.2, S.3 and S.4 report regression estimates of the models in Table 6 with alternate clustering of errors. Each table contains only a single column from Table 6. All regressions were run in Stata 16. Column 1, like Table 6 uses the `xtreg` command; however, rather than clustering errors on subjects, it reports SE estimates for clustering on session (xtset index is set session rather than subject ID). Column 2 uses the `reghdfe` command, which allows subject level effect to be "absorbed" into session effect. Column 3 uses the `cgmreg` command, which allows two-way clustering of errors. Two-way clustering accounts for covariance of error both by individual and session, but does not "nest" the former in the latter. Column 4 uses the `svy` preface, designed to organize data by primary sampling units. Across the various methods, the standard errors of the estimates for *Loss Framed* change only at the second decimal digit.

Table S.2 Estimated effect of loss-framed contracts on grids completed, using alternative variance estimators

	(1) xtreg	(2) reghdfe	(3) cgmreg	(4) svy:
Loss Framed	0.89	0.89	0.89	0.89
	[0.36,1.42]	[0.32,1.45]	[0.36,1.42]	[0.33,1.45]
Observations	536	536	536	536
Clustering	Session	Nested	Two-Way	Nested

95% CI in brackets, based on heteroskedastic-robust standard errors. All regressions also included dummy variables for order effects (=1 if started in loss frame), and for round effects (=1 if second round), whose estimated coefficients are suppressed for clarity.

Table S.3 Estimated effect of loss-framed contracts by contract preference on grids completed, using alternative variance estimators

	(1) xtreg	(2) reghdfe	(3) cgmreg	(4) svy:
Loss Framed	0.18	0.18	0.18	0.18
	[-0.48,0.84]	[-0.53,0.89]	[-0.48,0.84]	[-0.52,0.88]
Prefer Loss Frame	-2.60		-2.60	-2.60
	[-5.03,-0.18]		[-5.03,-0.18]	[-5.19,-0.01]
Prefer LF	3.28	3.28	3.28	3.28
& Loss Framed	[1.41,5.15]	[1.28,5.28]	[1.41,5.15]	[1.29,5.28]
Observations	536	536	536	536
Clustering	Session	Nested	Two-Way	Nested

95% CI in brackets, based on heteroskedastic-robust standard errors. All regressions also included dummy variables for order effects (=1 if started in loss frame), and for round effects (=1 if second round), whose estimated coefficients are suppressed for clarity.

Table S.4 Estimated effect of loss-framed contracts by contract preference (with indifference) on grids completed, using alternative variance estimators

	(1) xtreg	(2) reghdfe	(3) cgmreg	(4) svy:
Loss Framed	0.03	0.03	0.03	0.03
	[-0.82,0.87]	[-0.88,0.93]	[-0.82,0.87]	[-0.87,0.93]
Prefer Loss Frame	-2.71		-2.71	-2.71
	[-5.26,-0.15]		[-5.26,-0.15]	[-5.43,0.02]
Indifferent	-0.55		-0.55	-0.55
	[-2.65,1.56]		[-2.65,1.56]	[-2.79,1.70]
Prefer LF	3.46	3.46	3.46	3.46
& Loss Framed	[1.53,5.38]	[1.39,5.52]	[1.53,5.38]	[1.40,5.51]
Indifferent	0.91	0.91	0.91	0.91
& Loss Framed	[-0.62,2.45]	[-0.73,2.55]	[-0.62,2.45]	[-0.72,2.55]
Observations	536	536	536	536
Clustering	Session	Nested	Two-Way	Nested

95% CI in brackets, based on heteroskedastic-robust standard errors. All regressions also included dummy variables for order effects (=1 if started in loss frame), and for round effects (=1 if second round), whose estimated coefficients are suppressed for clarity.

S.2.3 Breaks

Table S.5 reports summary statistics for the number of breaks taken in the experiment.

Table S.5 Breaks by Frame

Frame	Round	Obs	Mean	SD	Min	Max
Gain Frame	Both	268	0.32	0.97	0	11
Gain Frame	Both	268	0.14	0.76	0	11
Gain Frame	1	135	0.34	1.20	0	11
Loss Frame	1	133	0.12	0.37	0	2
Gain Frame	2	133	0.29	0.66	0	3
Loss Frame	2	135	0.16	1.01	0	11

Table S.6 reports marginal effects from a probit regression on the likelihood of taking any breaks. In most rounds (465/536), workers took no breaks. In 49 rounds, they took one break and, in 13 rounds, they took two breaks. In only nine rounds did workers take more than two breaks. Given how rarely workers took multiple breaks, we define "taking a break" as a binary dependent variable, rather than a count variable, and estimate the effect of framing on breaks with a probit model. The estimated coefficients in the probit model imply that loss framing decreased the likelihood that a worker took a break. We cannot reject the null hypothesis that this effect is different between workers who preferred loss framing and workers who did not, but this subgroup hypothesis test has low statistical power given how rarely breaks were taken.

Table S.6 Estimated marginal effect of loss-framing on likelihood of taking a break

	(1) Impact of LF	(2) Impact by Preference	(3) Impact by Preference (w/ Indifference)
Loss Framed	-0.10	-0.10	-0.10
	[-0.16,-0.05]	[-0.17,-0.04]	[-0.17,-0.03]
Prefer Loss Frame		-0.02	-0.03
		[-0.11,0.07]	[-0.12,0.06]
Prefer LF & Loss Framed		0.01	0.01
		[-0.14,0.16]	[-0.15,0.16]
Indifferent			-0.07
			[-0.19,0.05]
Indifferent & Loss Framed			-0.04
			[-0.22,0.13]
Observations	536	536	536
Number of Subjects	268	268	268
Log Psuedolikelihood	-201	-201	-200

95% Confidence Interval in brackets, based on heteroskedastic-robust standard errors, clustered by worker. All regressions also included a dummy variable for order effects (=1 if started in loss frame) and for round effects (=1 if second round), whose estimated coefficients are suppressed for clarity.