Designing Contests for Data Science Competitions: Number of Stages and Prize Structures

Jialu Liu and Keehyung Kim

Production and Operations Management

E-Companion

Web Appendix A. Instruction for One-High Treatment

1. Introduction

This is an experiment in decision making. The instructions are simple - if you follow them carefully and make good decisions, you could earn a considerable amount of money which will be paid to you immediately following this experiment. What you earn partly depends on your decisions and partly on the decisions of others. Do not look at the decisions of others. You will be warned if you violate this rule. If you violate this rule twice, we will cancel the experiment immediately and your earnings will be $0.

The participants in this experiment will participate in a total of 15 decision rounds. At the start of each round, you will be randomly and anonymously assigned to a group consisting of 3 participants. This matching procedure will be repeated every round. We will use the computer to coordinate the experiment. The specific moves and decisions are described below.

2. Moves and Decisions

Please refer to the "Decision Cost Table" now. Your task in every round is to select a Decision Number, which ranges from 1 to 35. Associated with each Decision Number is a Decision Cost, which is listed on the same row in the next column. If you choose a higher Decision Number, you will have to pay a higher Decision Cost. This Decision Cost will be subtracted from your point earnings for that round.

After you have selected your Decision Number, the computer will generate your Random Number, which ranges from −15 to 15. Each number in this range has an equal chance of being drawn. Your Final Number is calculated as follows:

Final Number = Your Decision Number + Your Random Number

Based on Final Numbers of the three participants in your group, your Award will be determined in the following manner:

if your Final Number is *the **highest*** … **Your Award** is **8**	if your Final Number is ***not*** *the highest* … **Your Award** is **2**

The computer will then generate your **Point Earnings** in the following way:

Your Point Earnings = Your Award – Your Decision Cost

3. Cash Earnings

After the first round ends, we will repeat the same procedure for 14 more rounds. In each round, you are free to choose a different (or the same) Decision Number. Also, note that the computer will generate a Random Number separately in every round. We will then sum your point earnings across the 15 rounds. We will then multiply these point earnings by 0.064 and add $5.8 show-up fee to obtain your Cash Earnings.

Are there any questions?

Decision Cost Table

Decision Number	Decision Cost
1	0.006
2	0.022
3	0.050
4	0.089
5	0.139
6	0.200
7	0.272
8	0.356
9	0.450
10	0.556
11	0.672
12	0.800
13	0.939
14	1.089
15	1.250
16	1.422
17	1.606
18	1.800

Decision Number	Decision Cost
19	2.006
20	2.222
21	2.450
22	2.689
23	2.939
24	3.200
25	3.472
26	3.756
27	4.050
28	4.356
29	4.672
30	5.000
31	5.339
32	5.689
33	6.050
34	6.422
35	6.806

Web Appendix B. Range of Equation (14)

We need to derive the range of $\Delta = (e_{i,1}^* + e_{i,2}^*) - e_i^* = \frac{[6kv^2-(P_1-P_2)](P_1-P_2)}{144\,k^2v^3}$. Since $P_1 - P_2 > 0, 144k^2v^3 > 0$, whether Δ is positive or negative depends on the sign of $6kv^2 - (P_1 - P_2)$. Therefore, we can have two cases: $6kv^2 - (P_1 - P_2) > 0$ and $6kv^2 - (P_1 - P_2) \le 0$.

Case 1: $6kv^2 - (P_1 - P_2) > 0$:

In Case 1, $\Delta > 0$ because $6kv^2 - (P_1 - P_2) > 0, P_1 - P_2 > 0, 144k^2v^3 > 0$.

$$0 < \Delta = (e_{i,1}^* + e_{i,2}^*) - e_i^* = \frac{[6kv^2 - (P_1 - P_2)](P_1 - P_2)}{144\,k^2v^3} \le \frac{9k^2v^4}{144k^2v^3} = \frac{9v}{144}$$

The maximum of Δ is reached when $P_1 - P_2 = 3kv^2$.

Case 2: $6kv^2 - (P_1 - P_2) \le 0$:

In Case 2, $\Delta \le 0$ because $6kv^2 - (P_1 - P_2) \le 0, P_1 - P_2 > 0, 144k^2v^3 > 0$.

The Minimum value of Δ is obtained when $P_1 - P_2$ reaches maximum value. The difference between P_1 and P_2 is maximized under three prize structure: (a). winner-take-all prize structure: $P_1 = A, P_2 = 0, P_3 = 0$. (b). The last prize is zero and P_2 is the lowest possible value: $P_1 = A, P_2 = A - nkv^2, P_3 = 0$. (c). The last two prize are same: $P_1 = x + c, P_2 = P_3 = x$.

Case 2a. $P_1 = A, P_2 = 0, P_3 = 0, A \ge 6kv^2$:

The expected utility in *one-stage* contest EU_i should be positive, otherwise contestants will choose the lowest possible effort instead of equilibrium effort.

$$EU_i = \frac{A}{3} - ke_i^{*2} = \frac{A}{3} - k\left(\frac{A}{4kv}\right)^2 > 0$$

$\Rightarrow A < \frac{16}{3}kv^2$, which contradicts to $A \ge 6kv^2$.

Case 2b. $P_1 = A, P_2 = A - nkv^2, P_3 = 0, n \ge 6$:

The expected utility in *one-stage* contest EU_i should be positive, otherwise contestants will choose the lowest possible effort instead of equilibrium effort.

$$EU_i = \frac{A + A - nkv^2}{3} - ke_i^{*2} = \frac{2A - nkv^2}{3} - k\left(\frac{A}{4kv}\right)^2 > 0$$

$\Rightarrow k < 0$, which contradicts to $k > 0$.

Case 2c. $P_1 = x + c, P_2 = P_3 = x, c \ge 6kv^2$:

The expected utility in *one-stage* contest EU_i should be greater than the last prize x, otherwise contestants will choose the lowest possible effort instead of equilibrium effort.

$$EU_i = \frac{3x + c}{3} - ke_i^{*2} = \frac{3x + c}{3} - k\left(\frac{c}{4kv}\right)^2 > x$$

$\Rightarrow c < \frac{16}{3}kv^2$, which contradicts to $c \ge 6kv^2$.

In sum, Case 2 $(6kv^2 - (P_1 - P_2) \leq 0)$ does not exist. Therefore, the range of $\mathit{\Delta} = (e_{i,1}^* + e_{i,2}^*) - e_i^*$ is $(0, \frac{9v}{144}]$.

Web Appendix C. Experimental Results by the First Half Rounds and the Rest

Table C. Design and Summary Results of Experiment

(a) Rounds 1-8

Contest Structure		Base Model Prediction		Experiment result: average effort	
		Prize structure: High-spread	Prize Structure: Low-spread	Prize structure: High-spread	Prize Structure: Low-spread
One-stage contest	e_i	18	12	17.7 (11.8) N=312 t=-0.25, *p*=0.800	15.8 (9.4) N=336 t=3.13, *p*=0.002
Two-stage contest	$e_{i,1} + e_{i,2}$	18.6	12.7	35.6 (16.6) N=224 t=7.28, *p*<0.001	26.9 (12.5) N=224 t=8.54, *p*<0.001
	$e_{i,1}$	6.6	8.7	14.1 (10.0) N=336 t=5.77, *p*<0.001	14.0 (7.3) N=336 t=5.99, *p*<0.001
	$e_{i,2}$	12	4	18.5 (10.1) N=224 t=4.97, *p*<0.001	11.1 (8.8) N=224 t=5.92, *p*<0.001

Note: Numbers in parentheses are standard deviations. The t-statistics and *p*-values refer to the t-tests of the average effort from experimental result compared with the corresponding prediction of the base model. Standard errors are clustered at the participant's level. The reported results are for Rounds 1-8.

(b) Rounds 9-15

Contest Structure		Base Model Prediction		Experiment result: average effort	
		Prize structure: High-spread	Prize Structure: Low-spread	Prize structure: High-spread	Prize Structure: Low-spread
One-stage contest	e_i	18	12	15.2 (11.5) N=273 t=-2.12, *p*=0.035	14.0 (9.2) N=294 t=1.64, *p*=0.103
Two-stage contest	$e_{i,1} + e_{i,2}$	18.6	12.7	30.5 (14.8) N=196 t=5.90, *p*<0.001	22.1 (11.1) N=196 t=6.09, *p*<0.001
	$e_{i,1}$	6.6	8.7	11.5 (8.7) N=294 t=4.04, *p*<0.001	11.1 (6.7) N=294 t=2.58, *p*=0.010
	$e_{i,2}$	12	4	16.6 (9.4) N=196 t=3.65, *p*<0.001	9.2 (7.3) N=196 t=4.86, *p*<0.001

Note: Numbers in parentheses are standard deviations. The t-statistics and *p*-values refer to the t-tests of the average effort from experimental result compared with the corresponding prediction of the base model. Standard errors are clustered at the participant's level. The reported results are for Rounds 9-15.

Web Appendix D. Comparison between One-stage and Separate Stages of Two-stage for Main Experiment

In explaining why *two-stage* contest leads to a higher effort level, one of the alternative explanations was *mental accounting*. That is, contestants may consider each stage of *two-stage* contest as a separate independent *one-stage* contest. If contestants treat the two stages as two separate *one-stage* contests, we would observe that decisions in each stage of the *two-stage* contest are not significantly different from those of the *one-stage*. To test this possibility, we compare decisions between the *one-stage* treatment and each of the stages in the *two-stage* treatment. Statistics reported in Table D suggest that for high-spread prize structure, *one-stage* effort is significantly different from the effort in the first stage of *two-stage* contest (p=0.040). For low-spread prize structure, *one-stage* effort is significantly different from the effort in the second stage of *two-stage* contest (p=0.003). Since we observe clear differences between them, mental accounting is not likely a driving behavioral factor behind our experimental results.

Table D. Comparison between One-stage and Separate Stages of Two-stage for Main Experiment

Prize structure	Comparison	Experiment Result: Average Effort		t-statistics	p-values
		Contest structure: One-stage	Contest structure: Two-stage		
High-spread	e_i vs $e_{i,1}$	16.49	12.88	-2.06	0.040
	e_i vs $e_{i,2}$	16.49	17.61	0.65	0.516
Low-spread	e_i vs $e_{i,1}$	14.99	12.67	-1.62	0.106
	e_i vs $e_{i,2}$	14.99	10.21	-3.00	0.003

Note: t-tests are conducted between different contest structures under the same prize structure. Standard errors are clustered at the participant's level.

Web Appendix E. Validation Experiments

E.1. Experiment 2

Experiment 2 serves three purposes. First, in our main model, we assumed that contestants are risk neutral. In this validation experiment, we measure the risk preference of each contestant following the procedures of Holt and Laury (2002) and we rule out an alternative explanation by showing that risk-preference model cannot explain the observed behaviors. Second, we use a completely new set of parameters in this validation experiment to test whether our experimental results are robust and whether our findings are not specific to a certain choice of parameters. Third, we also test whether our estimated behavioral model (Section 5.4) can capture the true psychological drivers of contestants and whether it can explain contestants' behaviors under different experimental parameters. If our proposed behavioral model is robust, it should be able to predict contestants' behaviors under other experimental parameters as well. The detailed experimental design and results of validation experiment are reported below.

The experimental procedure of Experiment 2 is identical to that of Experiment 1. We used a new set of parameters as follows: $k = 1/200, v = 10, P_1 = 5.5, P_2 = 4, P_3 = 2.5$. To measure the risk preference of participants, when participants finished the experiment, they were given a questionnaire consisted of six lottery-choice questions (Table E1). Participants need to select one of the two lotteries – Option A (safer choice) or Option B (riskier choice) – for each of the six questions. The risk preference of contestants can be inferred from the number of safer choices (Option A) he or she makes before crossing over to riskier choices. Following previous literature (Lim et al., 2009), we adopt the constant relative risk aversion utility function $U(P) = \frac{P^\alpha}{\alpha}$ where α $(0 < \alpha \leq 1)$ to capture the degree of risk aversion where a lower α indicates greater aversion to risk. Using the responses from the lottery-choice questionnaires, we estimate the aggregate risk aversion parameter using probabilistic choice rule derived by Luce (1959). Specifically, the probability of choosing option A is $\Pr(A) = U_A^{\frac{1}{\mu}}/(U_A^{\frac{1}{\mu}} + U_B^{\frac{1}{\mu}})$, where U_A and U_B are the utility of Option A and Option B respectively. μ captures the insensitivity of choice probability (smaller values indicate greater sensitivity).

Table E1. Lottery Choice Question To Measure Risk Preference

Option A	Option B	Difference in Expected Value (A-B) ($) (not provided to participants)
40% chance of $20 and 60% chance of $16	40% chance of $38.5 and 60% chance of $1	1.60
50% chance of $20 and 50% chance of $16	50% chance of $38.5 and 50% chance of $1	-1.75
60% chance of $20 and 40% chance of $16	60% chance of $38.5 and 40% chance of $1	-5.10
70% chance of $20 and 30% chance of $16	70% chance of $38.5 and 30% chance of $1	-8.45
80% chance of $20 and 20% chance of $16	80% chance of $38.5 and 20% chance of $1	-11.80
90% chance of $20 and 10% chance of $16	90% chance of $38.5 and 10% chance of $1	-15.15

The maximum likelihood estimate of the aggregate risk aversion parameter is $\alpha = 0.418$ (p<0.001, log-likelihood: -259.45, μ=0.160). The estimated α falls within the range of 0.33 to 0.71, commonly observed risk-averse attitude among people as reported by Holt and Laury (2002). We also estimate the aggregate risk aversion parameter for *one-stage* contest ($\alpha = 0.457$, p<0.001) and *two-stage* contest ($\alpha = 0.380$, p<0.001) separately. Risk aversion parameters are not different across the two contest structures (W=0.601, p=0.438).

The left panel of Table E2 presents the predictions from the base model, the risk preference model, and the behavioral model. The prediction of the risk preference model is calculated based on the estimated aggregate risk aversion parameter ($\alpha = 0.418$). The behavioral model prediction is calculated based on equations (16), (19) and the solutions to equation (20) with the estimates from our main experiment ($\beta_o = 0.227, \beta_{T1} = 1.095, \beta_{T2} = 1.727, \theta_T = 0.006$ as shown in Table 7).

The experimental results of Experiment 2 are shown in the middle panel of Table E2. We found similar empirical regularities as observed in our main experiment. The total effort of *two-stage* contest is much higher (88% more) than that of *one-stage* contest although the base model suggests that they should be quite similar (with only 4% difference). Contestants significantly boost their effort provision in both the first stage and the second stage of *two-stage* contest. In terms of model prediction, the total efforts for *two-stage* contest are significantly different from base model predictions (t=8.05, p<0.001). Similarly, the experimental results are also significantly different from the risk preference model predictions (*one-stage*: t=7.76, p<0.001; *two-stage:* t=12.89, p<0.001). However, the total efforts from the experiment are not statistically different from the behavioral model predictions under both *one-stage* contest (t=-1.24, p=0.217) and *two-stage* contest (t=0.52, p=0.607). Furthermore, our behavioral model predicts that total effort of *two-stage* contest is 168% of the total effort in *one-stage* contest. Similarly, in experiment, the total effort of *two-stage* contest is indeed 188% of the total effort in *one-stage* contest. These show that either the base model or the risk preference model cannot fully explain the experimental results while our proposed behavioral model can predict contestants' behavior better in *two-stage* contest even with other parameters.

Table E2. Predicted and Actual Effort Decisions in Experiment 2

Contest Structure		Theory Prediction			Experiment Result	Test Against Theory Prediction		
		Base Model	Risk Preference	Behavioral Model		Base Model	Risk Preference	Behavioral Model
One-stage contest	e_i	15.0	6.85	16.70	15.3 (8.5)	t=0.32, p=0.751	t=7.76, p<0.001	t=-1.24, p=0.217
Two-stage contest	$e_{i,1} + e_{i,2}$	15.625	7.64	28.07	28.9 (13.0)	t=8.05, p<0.001	t=12.89, p<0.001	t=0.52, p=0.607
	$e_{i,1}$	10.625	5.62	15.42	15.3 (8.0)	t=4.41, p<0.001	t=9.13, p<0.001	t=0.11, p=0.913
	$e_{i,2}$	5.0	2.03	12.64	10.8 (8.2)	t=6.82, p<0.001	t=10.32, p<0.001	t=-2.18, p=0.030
Ratio of Two-stage to One-stage	$\frac{e_{i,1} + e_{i,2}}{e_i}$	**104%**	**115%**	**168%**	**188%**	-	-	

Note: Numbers in parentheses are standard deviations. The risk preference model is calculated using $\alpha = 0.418$. The t-statistics and p-values refer to the t-tests of the average effort from experimental results compared with the base model predictions and behavioral model predictions. Standard errors are clustered at the participant's level. The number of participants in *one-stage* contest and *two-stage* contest are 36 and 39.

E.2. Experiment 3

In our main experiment, we set the number of contestants to be $N = 3$. In Experiment 3, we vary the number of participants in a contest to check whether our findings can remain robust although the number of participants increases. Given the openness of the contest, it is important to test whether our main experimental results and behavioral model can be applied to the contests with more participants.

In this validation experiment, 5 contestants ($N = 5$) compete in a contest and the prize structure is $P_1 = 5, P_2 = 4, P_3 = 2, P_4 = 1, P_5 = 1$. We set $k = 1/150, v = 10$. In *two-stage* contest, 3 people will be shortlisted while 2 people will be eliminated after the first stage. Except for the abovementioned parameters, the experimental procedure of Experiment 3 is identical to our main experiment. We present the predictions from both the base model and the behavioral model on the left panel of Table E3. The behavioral model prediction is calculated with the estimates from our main experiment ($\beta_o = -0.227, \beta_{T1} = -1.095, \beta_{T2} = -1.727\,, \theta_T = -0.006$ as shown in Table 7).

The experimental results of the Experiment 3 are shown in the middle panel of Table E3. We found similar empirical regularities as observed in our main experiment. The total effort of *two-stage* contest is much higher (56% more) than that of *one-stage* contest although the base model suggests that they should be quite similar (with only 7% difference). Contestants significantly boost their effort provision in both the first stage and the second stage of *two-stage* contest. In terms of model prediction, the total efforts for *two-stage* contest are significantly different from base model predictions (t=4.77, p<0.001). However, the total efforts from the experiment are not statistically different from the behavioral model predictions under both *one-stage* contest (t=-1.58, p=0.116) and *two-stage* contest (t=-1.63 p=0.105). Furthermore, our behavioral model predicts that the total effort of *two-stage* contest is 151% of that in *one-stage* contest. Similarly, experimental results show the total effort of *two-stage* contest is indeed 156% of that in *one-stage* contest. Overall, these show that our behavioral model remains robust as the number of participants increases.

Table E3. Predicted and Actual Effort Decisions in Experiment 3

Contest Structure		Theory Prediction		Experiment	Test Against Theory Prediction	
		Base Model	Behavioral Model	Result	Base Model	Behavioral Model
One-stage contest	e_i	15.0	16.2	14.5 (6.2)	t=-0.44, p=0.658	t=-1.58, p=0.116
Two-stage contest	$e_{i,1} + e_{i,2}$	16.1	24.8	22.6 (7.6)	t=4.77, p<0.001	t=-1.63, p=0.105
	$e_{i,1}$	8.6	11.6	10.8 (5.3)	t=2.42, p=0.016	t=-0.85, p=0.394
	$e_{i,2}$	7.5	13.2	9.9 (5.5)	t=2.43, p=0.016	t=-3.40, p=0.001
Ratio of Two-stage to One-stage	$\frac{e_{i,1} + e_{i,2}}{e_i}$	**107%**	**151%**	**156%**	-	-

Note: Numbers in parentheses are standard deviations. The t-statistics and p-values refer to the t-tests of the average effort from experimental results compared with the base model predictions and behavioral model predictions. Standard errors are clustered at the participant's level. The number of participants in *one-stage* contest and *two-stage* contest are 20 and 20.

Similar to Section 5.4, we utilize the effort level decisions in validation experiments – Experiment 2 (*one-stage*: OV1, *two-stage*: TV1) and Experiment 3 (*one-stage*: OV2, *two-stage*: TV2) – to estimate the behavioral parameters $\beta_O, \beta_{T1}, \beta_{T2}, \theta_T$ via maximum likelihood method. The estimation results are displayed in Table E4 column (1). The estimates are largely consistent with those with our main experiment.

Table E4. Estimates of the Behavioral Model for Validation Experiment

	Parameter	(1) Behavioral model	(2) Base Model ($\boldsymbol{\beta} = \theta_T = 0$)
One-stage contest	β_O	-0.003	
		(0.037)	
	σ_{OV1}	8.486***	8.487***
		(0.250)	(0.250)
	σ_{OV2}	6.212***	6.222***
		(0.246)	(0.245)
Two-stage contest	β_{T1}	0.771***	
		(0.040)	
	β_{T2}	0.636***	
		(0.186)	
	θ_T	0.004***	
		(0.001)	
	$\sigma_{TV1,1}$	8.028***	9.255***
		(0.229)	(0.262)
	$\sigma_{TV1,2}$	8.193***	9.998***
		(0.287)	(0.347)
	$\sigma_{TV2,1}$	5.316***	5.705***
		(0.214)	(0.226)
	$\sigma_{TV2,2}$	5.580***	6.010***
		(0.290)	(0.307)
Log likelihood		-8329.5	-8538.2
LR Test			417.42***

Notes: $\boldsymbol{\beta}$ refers to β_O and β_T. LR test was against the behavioral model. Standard errors are shown in parentheses. ****p*<0.001

We found that $\beta_{T1} = 0.771$ is positive and statistically significant (*p*<0.001). The magnitude of β_{T1} is much larger than β_O (W=205.00, *p*<0.001), suggesting that being eliminated in *two-stage* contest is a more salient sign of "losing". Contestants also experience behind aversion in the second stage ($\beta_{T2} = 0.636$, *p*<0.001) and, at the same time, tend to stick with their first-stage effort in *two-stage* contest ($\theta_T = 0.004$, *p*<0.001). Moreover, we also estimate the base model in column (2) of Table E4 and the LR test rejects the base model (*p*<0.001). Together, these findings show that our proposed behavioral model is generalizable enough and robust with other parameters and more participants.

E.3. Experiment 4

Experiment 4 serves two purposes. First, we use another measurement of effort provision to check the validity of our results. Second, we vary the prize structure and allow the last prize to be $P_3 = 0$, which is a common practice in real data science contests (i.e., losing participants do not receive any monetary prize). This additional real-effort experiment hints that our experimental results based on the abstract model might be further generalized into a more realistic scenario where real physical efforts are invested.

In our main experiment as well as Experiment 2 and 3, we let participants choose a "Decision Number" and consider the "Decision Number" as the effort chosen by the participants. This method is called "stated effort" method in the literature of studying effort provision (Charness et al., 2018). With a stated-effort approach, subjects are presented with a list of decision options (i.e., effort choices) and their associated costs. The choice of "effort" involves a clear numerical cost, which influences the payoff of the subjects, as in a gift-exchange scenario (Fehr et al., 1993) or in a tournament (Bull et al., 1987). The advantage of the stated-effort approach is that there is no uncertainty regarding an individual's cost of effort. With full disclosure of the cost function, subjects can make an informed decision that maximizes their welfare. This makes us possible to test a theory via an abstract model in a lab and to identify empirical anomalies between the theory and the experiment. Furthermore, since we have full control over the relevant components of the model, we can test and compare both the base model and the behavioral model (section 5.4). One potential limitation of this approach is that simply selecting a number may not accurately represent the field environment due to abstraction of the reality. Thus, we conduct a real-effort experiment here to validate that our findings can be further generalized into a scenario where real physical efforts are invested.

In this new experiment, participants' effort is measured through real-effort tasks that assign them to particular observable tasks – in our case, slider tasks (Gill and Prowse, 2012). The slider task comprises of one screen with various sliders displayed on it. The number and position of the sliders on the screen remain the same regardless of the experimental subjects or decision rounds. When the subject first sees the screen with sliders tasks, all of the sliders are set to 0. The subject can move each slider to any integer position between 0 and 100 inclusively by using the mouse and the keyboard. It is possible to move and readjust each slider an unlimited number of times. Their efforts in the contest are measured as the number of sliders positioned at 50 at the end of the contest. No two sliders are perfectly aligned one under the other, making each slider equally difficult to position correctly. This stops the subject from being able to simply move the lower slider by replicating the position of the higher slider after positioning the higher slider at 50. Based on this setup, we used a 2 x 2 treatment design (*one-stage* vs. *two-stage, high-spread* vs. *low-spread*). The prize is given according to the rank of the number of sliders at position 50. In Experiment 4, we operationalize the *high-spread* prize structure as $P_1 = 12, P_2 = 0, P_3 = 0$ (*winner-takes-all*) and *low-spread* prize structure as $P_1 = 8, P_2 = 4, P_3 = 0$. All the other procedures of Experiment 4 are identical to

Experiment 1.

Table E5 summarizes the experiment results, which are largely consistent with the main findings from Experiment 1. The total effort of two-stage contest is much higher (138% more) than that of one-stage contest for both *high-spread* (*winner-takes-all*) and *low-spread* prize structure. Overall, while the *two-stage* contest produces a higher overall effort than the *one-stage* contest, and the high-spread prize structure of the two-stage contest leads to the highest overall effort.

Table E5. Result of Real-effort Experiment

Contest Structure		Experiment result: average effort Prize structure: High-spread	Prize Structure: Low-spread
One-stage contest	e_i	**19.9** (10.9) N=270	**15.1** (5.7) N=270
Two-stage contest	$e_{i,1} + e_{i,2}$	**47.3** (3.4) N=180	**36.0** (14.4) N=180
	$e_{i,1}$	28.8 (7.9) N=270	17.5 (11.6) N=270
	$e_{i,2}$	15.6 (5.8) N=180	14.1 (6.6) N=180
Ratio of Two-stage to One-stage	$\frac{e_{i,1} + e_{i,2}}{e_i}$	**238%**	**238%**

Note: Numbers in parentheses are standard deviations. The prize structure for *high-spread* contest is $P_1 = 12, P_2 = 0, P_3 = 0$. The prize structure for *low-spread* contest is $P_1 = 8, P_2 = 4, P_3 = 0$. Each treatment consists of 18 participants.

Reference

Bull, C., A. Schotter, K. Weigelt. 1987. Tournaments and piece rates: An experimental study. *Journal of political Economy*, 95(1),1-33.

Charness, G., U. Gneezy, A. Henderson. 2018. Experimental methods: Measuring effort in economics experiments. *Journal of Economic Behavior & Organization*, 149,74-87.

Fehr, E., G. Kirchsteiger, A. Riedl. 1993. Does fairness prevent market clearing? An experimental investigation. *The quarterly journal of economics*, 108(2),437-459.

Gill, D., V. Prowse. 2012. A structural analysis of disappointment aversion in a real effort competition. *American Economic Review*, 102(1),469-503.

Holt, C.A., S.K. Laury. 2002. Risk aversion and incentive effects. *The American economic review*, 92(5),1644-1655.

Lim, N., M.J. Ahearne, S.H. Ham. 2009. Designing sales contests: Does the prize structure matter? *Journal of Marketing Research*, 46(3),356-371.

Luce, R.D. 1959. *Individual choice behavior*. John Wiley & Sons, New York