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

Jialu Liu (corresponding author)
Antai College of Economics and Management, Shanghai Jiao Tong University, Shanghai, China
Email: jialuliu@sjtu.edu.cn | Phone: +86 (0)21 52302503

Keehyung Kim
School of Accounting and Finance, University of Waterloo, Waterloo, Ontario, Canada
Email: keehyung.kim@uwaterloo.ca | Phone: +1 (519) 888-4567

Abstract

Firms have been proactively holding data science competitions via online contest platforms to look for innovative solutions from the crowd. When firms are designing such competitions, a key question is: What should be a better contest design to motivate contestants to exert more effort? We model two commonly observed contest structures (*one-stage* and *two-stage*) and two widely adopted prize structures (*high-spread* and *low-spread*). We employ economic experiments to examine how contest design affects contestants' effort level. The results reject the base model with rationality assumption. We find that contestants exert significantly more effort in both the first stage and the second stage of the *two-stage* contest. Moreover, it is better to assign most prizes to the winner in the *two-stage* contest while it does not matter in *one-stage*. To explain the empirical regularities, we develop a behavioral economics model that captures contestants' psychological aversion to falling behind and continuous exertion of effort. Our findings demonstrate that it is important for contest organizers to account for the non-pecuniary factors that can influence contestants' behavior in designing a competition.

***Keywords*:** contest design, behind aversion, status quo, behavioral economics, experimental economics.
History: Received: August 2022; Accepted: August 2023 by Gary Bolton after two revisions.

1. Introduction

In recent years, firms have been proactively utilizing open competitions via online contest platforms (e.g., Kaggle, Tianchi, Biendata, Crowdanalytix, Tunedit) to look for innovative solutions from the crowd instead of developing their own solutions. These contest platforms let third parties compete against each other to come up with the most suitable solution for the firm's needs in exchange for a significant monetary prize. In 2018 alone, Kaggle had over 181,000 users participating in data science competitions on its platform.[1] Sometimes, the winner's prize goes as high as $1,000,000 in a single

[1] Data science competitions include problems in the area of machine learning, big data analytics, algorithm development, etc. The number of participants is reported in the article *Reviewing 2018 and Previewing 2019* posted in the Kaggle's official blog on Aug

competition.[2] This type of open competition offers two key benefits: First, firms could outsource their research and development (R&D) process and leverage both in-house expertise and outside talents to achieve a better solution. Second, a competition could serve as a way to strategically acquire talented people such as data scientists. Given the fixed prize budget allowed to design such competitions, a critical question that every contest organizer must confront is how to motivate contestants to exert more effort to come up with the best possible solution.

In most cases, contestants are generally ranked according to their score on a specific metric after submitting their solution and a higher prize is awarded to the contestant whose rank is higher. Sometimes, contest organizers implement an additional stage, hoping to find a better solution by letting contestants compete one more time. That is, some firms shortlist contestants first (i.e., elimination stage) and let only the shortlisted participants continue to participate in the next stage of the competition. In data science competitions, embedding an elimination stage can be easily implemented by limiting access to full data in the first stage. In specific, some data science competitions often involve two stages where a sample (or partial) data is offered in the first stage of the contest and, then, shortlisted participants from the first stage can continue with access to the full data in the second stage. A survey of the industry practices reveals that firms have indeed used various contest designs. To name a few, Google, Facebook, and McKinsey all adopted the simplest possible contest structure – that is, *one-stage* contest where all the contestants are competing against each other with the same data only one time.[3] On the other hand, Zillow and Microsoft implemented *two-stage* contest where the contestants compete to solve one problem in the first stage and the shortlisted contestants keep solving the same problem using more data in the second stage.[4] In fact, *two-stage* contests are becoming increasingly popular as most of the data science competitions held in Tianchi, the biggest data science competition platform in China, are adopting two stages.[5] In this type of *two-stage* contest structure where the task is identical across both stages, contestants keep building their second stage solution on top of their first-stage solution. Thus, when contestants are evaluated, their final performance is seen as the combination of their first-stage performance and the additional performance

12, 2019. See: https://web.archive.org/web/20190812203743/http://blog.kaggle.com/2019/01/18/reviewing-2018-and-previewing-2019/

[2] The total prize of the Zillow's Home Value Prediction Competition was $1,150,000. See: https://www.kaggle.com/c/zillow-prize-1/overview/prizes

[3] Google hosted a Kaggle competition to build an algorithm that detects objects automatically. See: https://www.kaggle.com/c/google-ai-open-images-object-detection-track/overview/prizes Facebook hosted a Kaggle competition to predict a user's choice of a hotel. See: https://www.kaggle.com/c/facebook-v-predicting-check-ins/overview/description McKinsey hosted a competition on Analytic Vidhya to solve its daily challenge. See: https://datahack.analyticsvidhya.com/contest/mckinsey-analytics-online-hackathon-4/

[4] Zillow hosted a Kaggle competition to estimate home values. See: https://www.kaggle.com/c/zillow-prize-1/overview/ Microsoft hosted a competition on Biendata to extract scholars' profile from text. See: https://biendata.com/competition/scholar/timeprize/

[5] Tianchi (https://tianchi.aliyun.com/competition/gameList/activeList), owned by Alibaba, held more than 200 data science competitions and has at least 120,000 participants as of 2018, according to Wired UK.

obtained in the second stage.

While there could be potentially various ways in designing a multi-stage contest in a traditional setting, the *two-stage* contest studied in this paper follows the unique structure commonly adopted by firms holding data science competitions. That is, it is different from other traditional types of multi-stage contests designed for sports tournaments (Sheremeta 2010a) or political campaigns (Sheremeta 2010b). For sports tournaments (e.g., FIFA World Cup) (Sheremeta 2010a), the performance achieved in previous stages does not increase the chance of winning in the next stages (because each football match is independent), whereas in our context, the previous performance affects the probability of winning the subsequent contest. For example, in a data science competition with two stages, if the first-stage performance of a contestant's solution is exceptional with partial data, the contestant can win even without any improvement in the next stage because the solution from the first stage can continue to be outperforming with full data. In terms of political campaigns (e.g., US presidential election) (Sheremeta 2010b), each party first selects one candidate in the first stage, and then they compete with competitors from other parties in the final stage (e.g., general election). In this case, participants are first divided into subgroups and the winners from subgroups are competing again. On the contrary, in our context, contestants are simply shortlisted based on their first-stage performance and those compete again in the next stage of the *two-stage* contest. Thus, the *two-stage* contest design in our paper reflects the unique characteristics of data science competitions where contestants are asked to solve the same (or similar) problem across stages.

In addition to the contest structure (i.e., the number of stages), another variation frequently observed in designing a contest is prize structure (Lim et al. 2009). Dato and FIND adopted *high-spread* prize structure, in which the contestant with the highest performance gets almost all the prize.[6] The most extreme case of *high-spread* prize structure is a *winner-takes-all*, in which there is only one winner in the contest. Other companies such as Quora, Alibaba, and Youku used *low-spread* prize structure – that is, there are multiple winners and the rank-ordered prizes awarded to them are more equally distributed.[7] In this case, the prize difference between two winners consecutively ranked next to each other is relatively small. Given the varying potential designs of the contest structure and the prize structure that firms can

[6] Dato hosted a Kaggle competition to predict which web pages served by StumbleUpon are sponsored. It provided $5,000 for the first place and others would earn nothing. See: https://www.kaggle.com/c/dato-native/overview/prizes. FIND Technologies Inc. hosted a competition on Tunedit for an algorithm that identifies and categorizes data of electromagnetic signatures into 3 classes of substances, at an accuracy of 95% or higher. The top winner would receive $40,000 (Canadian dollars) and the remaining five winners would get $1,000. See: http://tunedit.org/challenge/material-classification

[7] Quora hosted a Kaggle competition to detect toxic content. The first prize, second prize and third prize are $12,000, $8,000 and $5,000 respectively. See: https://www.kaggle.com/c/quora-insincere-questions-classification/overview/prizes Alibaba hosted a competition on Tianchi. The first, second and third prize are RMB 80,000, RMB 50,000 and RMB 30,000. See: https://tianchi.aliyun.com/competition/entrance/231665/introduction Youku hosted a competition on Tianchi. The prize structure is RMB100,000, RMB 60,000, and RMB 40,000 for the first, second and third rank. See: https://tianchi.aliyun.com/competition/entrance/231711/introduction

adopt in practice, we want to answer the following questions: In holding a data science competition, does the contest structure matter? Specifically, should the contest include one stage only or two stages? Additionally, given the contest structure, does the prize structure (*high-spread* vs. *low-spread*) matter?

This paper investigates multiple contest designs which are commonly observed in data science competitions and examines how a contest organizer can maximize the effort level exerted by contestants given a fixed prize budget. To achieve our goal, we propose a base model that abstracts a data science competition where contestants are competing for fixed prizes. To observe and compare behaviors of contestants across potential contest designs, we employ a 2×2 experimental design by manipulating two treatment factors. The first factor that we manipulate is the contest structure where there could be one or two stages in a contest (*one-stage* vs. *two-stage*). In *one-stage* treatments, contestants compete for one stage only and then their award is determined based on the performance from the single stage. In *two-stage* treatments, all contestants decide their first-stage effort level in the first stage, and then only shortlisted contestants decide their additional effort level in the second stage. Their award is determined based on the total performance, that is, the sum of the first-stage performance and the second-stage performance. For the second factor, we manipulate the prize structure where prizes could be allocated mostly to a single winner or more equally distributed among multiple winners (*high-spread* vs. *low-spread*). For *high-spread* treatments, the difference between the winner's prize and the others' is large whereas, in *low-spread* treatments, such difference is relatively small.

Our paper contributes to the growing stream of Operation Management (OM) literature studying contest design. Many previous studies have focused on a theoretical model to look for the optimal contest design in constant-sum competitions (Alpern and Howard 2017), multiple parallel contests (Körpeoğlu et al. 2022), and internal contests (Nittala et al. 2022). Regarding contest features, prize structure (Bimpikis et al. 2019, Korpeoglu et al. 2021, Nittala et al. 2022), information disclosure (Bimpikis et al. 2019), entry policies (Ales et al. 2021), and contest duration (Korpeoglu et al. 2021) have investigated. Other previous studies have utilized secondary data to investigate the impact of a particular contest feature such as contestants' prior experience (Menon et al. 2020), contests' problem specification (Jiang et al. 2021), and information disclosure (Wooten 2022) on contest outcomes. Extending previous findings, our research leverages an incentive-aligned laboratory experiment (e.g., Davis 2015, Davis and Hyndman 2018, Davis et al. 2021) to study the optimal contest design, which allows us to overcome several major challenges in empirically examining the effect of the contest design. First, we can make exogenous variations in the contest structure and the prize structure across the treatments. Then, we are able to make random assignments of participants into different treatments, which cannot be easily done in the field. Second, the effort level invested by a contestant is a strategic variable in a contest and it can vary with other factors such as whether there are superstar competitors (Zhang et al. 2019) and whether contest organizers provide exemplars (Koh 2019). We can control these factors and focus on the impact of treatment factors in a controlled setting. Third, in real life, it could be that a contestant spends a small amount of effort, but the solution designed by the contestant achieves good performance due to luck. Given our experimental setup, we can accurately measure the effort provision as well as the performance in the experiment whereas we can only observe the performance (but, not the effort provision) in the field. This is important because the contest organizer cannot alter the ability or the luck of the contestants, but she can only motivate the effort provision of the contestants in order to achieve a better solution. Thus, a clear measurement on the effort level is an inevitable step to test which factors influence the effort provision in designing a competition, which can be more effectively done through a lab experiment.

The results of our Experiment 1 show the followings: First, we found that contestants' effort provision in *one-stage* treatments is overall similar to (slightly higher than) the prediction under *high-spread* (*low-spread*) condition when the prediction is from the base model assuming that players are rational. However, in the case of *two-stage* treatments, contestants significantly boost their first-stage effort in order to be shortlisted and, then, they continue to over-exert their effort in the second stage. As a result, in the *two-stage* treatments, contestants tend to invest significantly higher effort than they do in *one-stage*. In addition to the contest structure, we also report the impact of the prize structure. While the prize structure does not affect the effort provision in the *one-stage* treatments, contestants in the *two-stage* treatments are shown to spend more effort under *high-spread* prize structure. These empirical anomalies cannot be explained by the base model alone, which suggests the need to account for behavioral components.

Furthermore, our work is related to the extensive literature in behavioral OM, which challenges the main underlying assumption of fully rational profit-maximizing decision makers in standard OM models. In behavioral OM, incorporating behavioral or psychological factors into the player's utility has been shown to explain decision-making behaviors better in various contexts including, but not limited to, newsvendor problems (Schweitzer and Cachon 2000, Bolton and Katok 2008, Becker-Peth et al. 2020) and wholesale supply chain (Davis et al. 2014, Davis 2015). Our paper also contributes to this stream by providing a formal explanation of the behavior observed in the experiment. We develop a behavioral model that generalizes the base model by capturing the psychological utilities of the subjects and we econometrically estimate it using our experimental data (e.g., Davis 2015, Davis and Hyndman 2018). We show that our generalized model tracks behaviors much better than the base model and other nested models. First, contestants exhibit a psychological aversion to being eliminated early. Specifically, having a second stage makes the separation of "winning" and "losing" more salient compared to the one-stage contest. Being eliminated in the first stage of a *two-stage* contest clearly signifies losing. In order to avoid "being left behind," also conceptualized as "behind aversion" (Roels and Su 2014), contestants exert much more effort than expected in the first stage, which explains our first empirical anomaly. Second, contestants are biased toward their previous decisions (the first-stage effort) and continue exerting a significant amount of effort in the second stage. Previous literature suggests that individuals tend to maintain their previous decisions and disproportionately stick with the status quo (Samuelson and Zeckhauser 1988). This provides contestants with a strong impetus to keep exerting effort in the second stage, which explains our second empirical anomaly.

We also provide further validation of our proposed behavioral model by conducting an additional experiment with a different set of experimental parameters and other control factors that may affect results. In Experiment 2, we incorporate the risk preference of the contestants, and in Experiment 3, we vary the number of participants in a contest. Overall, our parameterized behavioral model with estimates from Experiment 1 is able to predict the results of the new experiments quite well. Moreover, in Experiment 4, we conduct an additional real-effort experiment to show that our experimental results based on the abstract model can be further generalized into a more realistic scenario where real physical efforts are invested. Overall, we demonstrate that it is important for the contest organizer to be cognizant of the nonpecuniary drivers of contestants in designing data science competitions.

The rest of the paper is organized as follows: Section 2 reviews related literature and explains the potential contribution of our paper. We introduce the models for two different contest structures (*one-stage* and *two-stage*) and our experimental design in Section 3. In Section 4, we describe our experimental results and highlight the main empirical regularities. To explain these, in Section 5, we develop a behavioral model that captures contestants' psychological aversion to falling behind and continuous exertion of effort and estimate the behavioral model using the experimental data. Section 6 describes three additional experiments that further validate the robustness of our behavioral model. Lastly, Section 7 concludes.

2. Related literature

2.1. Tournament Theory and Contest Design

Our work builds on the literature of tournament theory. Charness and Kuhn (2011) and Dechenaux et al. (2015) provide a comprehensive literature review on tournament theory. Previous literature on tournament theory considers questions on selection into tournaments (Lazear and Rosen 1981), sabotage (Lazear 1989), collusion (Harbring and Irlenbusch 2003), etc. Our paper adds to one important field in the tournament theory – that is, *contest design*. Kalra and Shi (2001) and Moldovanu and Sela (2001) used theoretical models to demonstrate the optimal prize structure in a single-stage contest based on the assumption of rational players. Their model shows that the optimal contest should have only one winner (i.e., *winner-take-all*). Later, their findings are empirically challenged by Lim et al. (2009), which experimentally demonstrates that the prize structure has little impact on contestants' decisions. Our study extends this question by taking a multi-stage contest design into consideration. We also vary the prize structure to see how the contest structure's effect on the contestants' decision interacts with the prize structure.

Furthermore, researchers have also studied other important aspects of contest design. One aspect is the role of information disclosure in an optimal contest design. Bimpikis et al. (2019) analyze whether and when the contest designer should disclose information regarding the competitors' progress to maximize the designer's expected payoff. Wooten (2022) discovers that leaps improve overall contest performance and boost participation rates for complex contests. Other important aspects of contest design include entry policies (Ales et al. 2021), contest duration (Korpeoglu et al. 2021), contestants' prior experience (Menon et al. 2020), and contests' problem specification (Jiang et al. 2021). Our paper adds to this stream of OM literature by comparing contestants' behaviors across multiple contest designs, developing a behavioral model that could explain such contestants' behaviors, estimating it using the experimental data, and validating it with additional experiments.

2.2. Multi-stage Contest Design

Our work is related to the literature about multi-stage tournament design (See Dechenaux et al. 2015 for a comprehensive review). Multi-stage contest design is rooted from two models – rent-seeking contests (Tullock 2001) and rank-order contests (Lazear and Rosen 1981). In rent-seeking contests, contest designers wish to minimize efforts spent by contestants because these efforts are regarded as social waste (Tullock 2001, Dechenaux et al. 2015), which is usually applied in political competitions or competitions for government subsidy (and public goods). Examples focusing on multi-stage rent-seeking contests are listed in Table 1. In contrast to this, in rank-order contests, efforts from contestants are valuable and contest organizers wish to maximize the efforts. Our contest design lies in the second category (rank-order contest) where contest organizers wish to maximize the effort exerted by contestants. Furthermore, for a typical data science competition, the problem that contestants need to solve is identical across stages and thus the performances in different stages are considered dependent because the first-stage solution can be used in the next stage. This design is unique and different from multi-stage rank-order contests studied previously (e.g., Altmann et al. 2012, Delfgaauw et al. 2015) which assume that the performance in each stage is independent. Thus, our paper is the first to study a unique context of rank-order contests where the total prize is determined based on the total efforts across stages.

Overall, the multi-stage design in our paper captures the unique characteristics of data science competitions which previous papers have not studied. Table 1 summarizes the literature about multi-stage tournaments and presents the differences between prior works and our paper (last row). First, columns (1)-(2) of Table 1 list the model choice (rent-seeking contests vs. rank-order contests) of the corresponding paper. Our paper leverages the rank-order contest because prize is awarded based on the rank of the contestant in data science competitions. Second, typical two-stage data science competitions shortlist top-performing participants after the first round while it eliminates others from further participation (column

(3)). Third, to reflect the unique context of solving the same questions across stages, the final prize should be determined by the total effort instead of the effort in the second stage alone (column (4)). Fourth, contestants in our contest are competing against all participants at the same time instead of competing within a subgroup (column (5)). Previous literature notices that competing against all contestants often yields different results from competing within subgroups (Kalra and Shi 2001). Our paper is unique in modeling the data science competitions by including all the four features mentioned above (columns (2)-(5)). Moreover, our paper also provides an empirical comparison between *one-stage* contest and *two-stage* contest while previous studies often focused on the *two-stage* contest design alone (column (6)). To ease comparison between *one-stage* contest and *two-stage* contest, we let the total prize to be the same across different contest structures (column (7)).

Table 1. Summary of Previous Studies in Multi-stage Contests and Our Paper

Paper	Rent-seeking Contests	Rank-order Contests	Elimination	Prize determined by total efforts	No subgroup	Comparison between one-stage and two-stage contest	Total prize same between one-stage and two-stage contest
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Schmitt et al. (2004)	✓			✓	✓	✓	
Parco et al. (2005)	✓		✓				
Amaldoss and Rapoport (2009)	✓		✓				
Stracke et al. (2015)	✓		✓			✓	✓
Sheremeta (2010a)	✓		✓	✓			
Sheremeta (2010b)	✓		✓			✓	✓
Fu and Lu (2012)	✓		✓	✓	✓	✓	✓
Altmann et al. (2012)		✓	✓		✓	✓	
Delfgaauw et al. (2015)		✓	✓				
Mago and Sheremeta (2019)	✓				✓		
Deutscher et al. (2019)		✓					
Our Paper		✓	✓	✓	✓	✓	✓

3. Model Overview

3.1. Models of Contest

We first introduce a base model of a *one-stage* contest and, then, subsequently show a base model of a *two-stage* contest.[8] In both models, there are three contestants competing against each other for prizes. Here, for each model, we also provide the theoretical equilibrium based on the assumption of players being rational.[9] Note that, to cleanly identify the effect of contest structure and prize structure in experiments, we start by proposing the simplest possible model. Later in our validation experiments (and Appendix E), we further extend by incorporating the risk preference of the contestants and varying the number of participants in a contest. Our models allow us to experimentally examine the impact of multiple factors (contest and

[8] We use the term "base model" throughout the paper to refer to the models presented in this section (3.1).

[9] Players are assumed to be risk-neutral in our base model. Our validation experiment (Appendix E.1) incorporates the risk preference of the contestants.

prize structure) in designing a contest and further observe if a certain design motivates contestants to exert more effort.

3.1.1. A Model of One-stage Contest

Consider a contest that consists of $N = 3$ contestants competing for prizes based on their performances. The prize structure of the contest is given by $P_1 \geq P_2 \geq P_3$, where P_1 is the monetary value received by the contestant with the highest rank, and so on. The contestant i individually decides to spend effort e_i and achieve the performance y_i according to the following function:

$$y_i = e_i + \epsilon_i \tag{1}$$

where ϵ_i is a random shock faced by the contestant i, which reflects the uncertainty and other unanticipated factors outside the contestant's control on performance. This assumes that, given the higher effort, a higher performance is expected in general. Shocks ϵ_i are assumed to be drawn independently from the uniform distribution $unif(-v, v)$. In real life, because of the shock, the contest organizer cannot observe the effort e_i expanded by the contestant to achieve a given performance y_i whereas we exactly observe how much effort the contestant spent in an experimental setting. Lastly, there is a cost associated with the effort as specified by the cost function $c(e_i) = ke_i^2$, which is strictly increasing and convex. That is, higher the effort, higher the chance to win but also higher the cost. Therefore, the payoff function of the contestant is:

$$\pi_i = P_i - c(e_i) \tag{2}$$

where P_i is the monetary prize given to the contestant i based on his rank. Thus, to maximize the payoff, a contestant should evaluate the trade-off between winning a higher prize and incurring a higher cost by expanding effort.

Equilibrium of One-stage Contest. Following the definition of "rational players" in previous literature (Mullainathan and Thaler 2000, Kalra and Shi 2001, Lim 2010), rational contestants are assumed to care about the monetary payoff they receive. Based on the payoff function presented on equation (2), each contestant decides his effort level e_i^* that maximizes his expected payoff:

$$e_i^* = \underset{e_i}{\operatorname{argmax}}\, E\pi_i = \underset{e_i}{\operatorname{argmax}} \sum_r Prob(P_i = P_r) \times P_r - c(e_i) \tag{3}$$

where $Prob(P_i = P_r)$ denotes the probability that the contestant i's performance y_i ranks the r^{th} position among the three contestants and he will be awarded monetary prize P_r. Ranking the r^{th} position happens when the contestant's performance y_i is lower than the performance of other $r - 1$ contestants and higher than the other $N - r$ contestants. Considering that all contestants follow the same decision-making process, the probability of ranking the r^{th} position is (Kalra and Shi 2001):

$$Prob(P_i = P_r) = \int_{-v}^{v} \binom{N-1}{r-1} [1 - F(e_i - e_i^* + \epsilon_i)]^{r-1} F^{N-r}(e_i - e_i^* + \epsilon_i) f(\epsilon_i) d\epsilon_i \tag{4}$$

where $F(\cdot)$ and $f(\cdot)$ are the CDF and PDF of random shock ϵ_i, which is drawn from the uniform distribution.

To maximize the expected payoff ($E\pi_i$) in equation (3), we solve its first order condition for the optimal effort level:

$$\sum_{r=1}^{N} \frac{\partial Prob(P_i = P_r)}{\partial e_i}(e_i = e_i^*) \times P_r - c'(e_i^*) = 0 \tag{5}$$

According to Kalra and Shi (2001) and Orrison et al. (2004), the marginal probability of achieving rank r given the effort $e_i = e_i^*$ would be:

$$\frac{\partial Prob(P_i = P_r)}{\partial e_i}(e_i^*) = \begin{cases} \frac{1}{2v}, & if\ r = 1 \\ 0, & if\ r = 2 \\ -\frac{1}{2v}, & if\ r = 3 \end{cases} \tag{6}$$

Substituting equation (6) into equation (5), we can solve the symmetric pure-strategy Nash equilibrium of the optimal effort for a given prize structure of *one-stage* contest as follows:

$$e_i^* = \frac{P_1 - P_3}{4kv} \tag{7}$$

As shown on equation (7), the equilibrium effort by contestants is determined by the difference between the first prize and the last prize and it is irrelevant to the second prize.[10] This suggests that, in order to obtain maximum performance from contestants, the contest organizer should adopt a *winner-take-all* prize structure – that is, except the first rank, other contestants should be awarded the lowest possible prize.

3.1.2. A Model of Two-stage Contest

We now extend the model of *one-stage* contest to include one additional stage. *Two-stage* contest is different from the *one-stage* in that: (1) *Two-stage* contest consists of two decision stages for contestants; (2) Only two contestants with higher first-stage performance among the three contestants can participate in the second stage. The contestant with the lowest first-stage performance cannot proceed to the second stage; (3) The prizes given to the contestants are determined by the sum of the performance in both stages.

The whole procedure of *two-stage* contest is as follows. Three participants compete against each other in the first stage. The contestant whose first-stage performance is the lowest will get the prize P_3 and he is eliminated from further competition. The other two contestants (whose first-stage performance is not the lowest) will continue to participate in the second stage and compete against each other. The participant whose sum of the first-stage performance and the second-stage performance is the highest will win the award P_1 and the other contestant gets the award P_2. This *two-stage* contest is modeled as follows. Contestant $i \in \{1, 2, 3\}$ who participates in stage $k \in \{1, 2\}$ achieves performance $y_{i,k}$ according to:

$$y_{i,k} = e_{i,k} + \epsilon_{i,k} \tag{8}$$

[10] This is consistent with the theoretical prediction of Kalra and Shi (2001) and Lim (2010).

where $e_{i,k}$ is the effort exerted by contestant i in stage k, and $\epsilon_{i,k}$ is a random shock faced by contestant i in stage k. Shock $\epsilon_{i,k}$ is drawn independently for each contestant in each stage from the uniform distribution $unif(-v, v)$. If the contestant cannot participate in the second stage, his $e_{i,2} = 0$ and $y_{i,2} = 0$. The contestant's cost function is $c(e_{i,1}, e_{i,2}) = ke_{i,1}^2 + ke_{i,2}^2$. Therefore, the payoff function of the contestant is:

$$\pi_i = P_i - c(e_{i,1}, e_{i,2}) \tag{9}$$

where P_i is the monetary prize given based on the contestant i's rank.

Equilibrium of Two-stage Contest. In this section, we will derive the closed form solution for the *two-stage* contest. We use backward induction to solve the problem. In the second stage, two participants choose to exert effort $e_{i,2}$ to get total performance $y_i = y_{i,1} + y_{i,2} = y_{i,1} + e_{i,2} + \epsilon_{i,2}$. We define $\Delta\epsilon_k = \epsilon_{j,k} - \epsilon_{i,k}$ as the difference between the random shocks of the two contestants who are allowed to participate in the second stage ($k = 2$) with the probability density function $g_{\Delta\epsilon_k}$, which is a triangular distribution because shock $\epsilon_{i,k}$ is drawn from the uniform distribution. The expected payoff for contestant i in the second stage is as follows:

$$E\pi_{i,2} = Prob(y_i > y_j) \times P_1 + Prob(y_j > y_i) \times P_2 - c(e_{i,2}) \tag{10}$$

where $Prob(y_i > y_j)$ is the probability that contestant i's total performance is higher than the other competing contestant j's.

The first order condition of the second-stage expected payoff $E\pi_{i,2}$ in equation (10) would be:

$$(P_1 - P_2)\mathbb{E}_{\epsilon_1}[g_{\Delta\epsilon_2}(y_{i,1} + e_{i,2} - y_{j,2} - e_{j,2})] - c'(e_{i,2}) = 0 \tag{11}$$

Because $g_{\Delta\epsilon}$ is symmetric around zero and contestants all play equilibrium efforts $e_{i,1}^*$ and $e_{i,2}^*$ in the first stage and the second, the unique second-stage subgame perfect Nash equilibrium effort would be:[11]

$$e_{i,2}^* = \frac{(P_1 - P_2)}{2k}\mathbb{E}_{\epsilon_1}[g_{\Delta\epsilon_2}(\Delta\epsilon_1)] = \frac{(P_1 - P_2)}{6kv} \tag{12}$$

Plugging equation (12) into equation (10), we can solve the expected payoff of entering the second stage $E\pi_{i,2} = \frac{P_1+P_2}{2} - c(e_{i,2}^*)$. Then, it reduces to a *one-stage* contest with $N = 3$ participants and the prize structure $P_1 = P_2 = E\pi_{i,2} \geq P_3$. Substituting this prize structure into equation (7), we can derive the first-stage equilibrium effort as follows:

$$e_{i,1}^* = \frac{E\pi_{i,2} - P_3}{4kv} = \frac{18kv^2P_1 + 18kv^2P_2 - P_1^2 + 2P_1P_2 - P_2^2 - 36kv^2P_3}{144k^2v^3} \tag{13}$$

In order to maximize the equilibrium effort by contestants in *two-stage* contest, contest organizers should make the prize spread as big as possible because of the following reasons: First, equation (12) suggests that, to maximize the second-stage equilibrium effort $e_{i,2}^*$, the difference between P_1 and P_2 should

[11] $\mathbb{E}_{\epsilon_1}[g_{\Delta\epsilon_2}(\Delta\epsilon_1)] = \int_{-2v}^{2v} g_{\Delta\epsilon_2}(z)\, g_{\Delta\epsilon_1}(z)dz = \int_{-2v}^{2v} g_{\Delta\epsilon_1}(z)^2\, dz = \frac{1}{3v}$

be as big as possible. Second, based on equation (13), the first-stage equilibrium effort $e_{i,1}^*$ can be maximized when the difference between the second-stage's expected award $(P_1 + P_2)/2$ and the last prize P_3 is big. Based on the two models presented here, we propose our experimental design and parameterize our models next. Then, we subsequently present our model predictions in Section 3.4.

3.2. Experimental Design

Our main experiment employs a 2×2 factorial between-subject design to examine the following two questions: First, how do different contest structures affect effort provision? Second, how do prize structures influence effort provision? We elaborate on our treatments below.

The first factor that we study is contest structure (*one-stage* contest vs. *two-stage* contest) where we manipulate the number of stages in the contest. In *one-stage* treatments, the contest consists of only one stage. Contestants' ranks and prizes are determined by their performance based on a single decision task, which is a function of their effort and random shock. In *two-stage* treatments, the contest consists of two stages. The first-stage performance will decide whether or not the contestant is allowed to participate in the second stage. The total performance determines the prize earned. That is, both the first-stage decision and the second-stage decision can affect the final rank and the final prize. The second treatment variable that we manipulate is prize structure (*high-spread* vs. *low-spread*). In *high-spread* treatments, the spread of prizes is wide. We operationalize this by setting the top prize as 8 and both the second and third prizes as 2. Thus, the difference between the highest prize and the lowest is as high as 6. In *low-spread* treatments, the difference between two consecutive prizes is equal to 2. That is, the first prize is 6, the second prize is 4, and the third prize is 2. In this way, we are able to fix the total budget for prizes in both *high-spread* and *low-spread* treatments to be the same at 12. This design yields four treatments in Table 2.

Table 2. Summary of Treatments

#	Treatment Label	Treatment Abbreviation	Contest Structure	Prize Structure
1	One-High	OH	One-stage contest	$P_1 = 8, P_2 = 2, P_3 = 2$
2	Two-High	TH	Two-stage contest	$P_1 = 8, P_2 = 2, P_3 = 2$
3	One-Low	OL	One-stage contest	$P_1 = 6, P_2 = 4, P_3 = 2$
4	Two-Low	TL	Two-stage contest	$P_1 = 6, P_2 = 4, P_3 = 2$

One-High Treatment (OH). In this treatment, the contest consists of only one stage and the prize spread is high. Here, we detail how the parameters for the *high-spread* prize structure are chosen. Note that, the model equilibrium shown in equation (7) suggests that, given a fixed budget, a *winner-take-all* prize structure can maximize effort provision from contestants. Thus, the first rank should take the most prizes while the other contestants should be awarded the lowest possible prize. In our experiment, we set the total budget to be 12. Ideally, prize structure of $P_1 = 12, P_2 = 0, P_3 = 0$ would lead to the highest possible effort in theory.[12] However, in this case, if contestants exert effort greater than 0, two-thirds of the

[12] In Experiment 4 (Appendix E.3), we conduct an additional experiment with *winner-takes-all* prize structure.

contestants' payoff would be negative under the equilibrium (equation (7)) because such effort bears a significant cost. Therefore, we set prize structure $P_1 = 8, P_2 = 2, P_3 = 2$ to make sure that we mimic the scenario where prize structure is as spread as we can and, at the same time, contestants would obtain positive payoffs from our experiment.[13] We set $v = 15$ and $k = 1/180$ for the shocks and cost parameters.

Two-High Treatment (TH). The treatment consists of two stages in the contest. In the first stage, the last prize is given to the last performer and the other two contestants proceed to the second stage. In the second stage, top 2 prizes are given according to the total performance (i.e., the sum of the first-stage performance and the second-stage performance). The prize structure is the same as One-High (OH).

One-Low Treatment (OL). The treatment is identical to One-High above with one exception: the prize structure. Here, the difference between consecutive prizes is smaller and prizes are more equally distributed. We operationalize this *low-spread* prize structure by setting $P_1 = 6, P_2 = 4, P_3 = 2$ as mentioned above.

Two-Low Treatment (TL). This treatment is identical to Two-High above except that the prize structure is the same as One-Low: $P_1 = 6, P_2 = 4, P_3 = 2$.

The parameter setup presented here is for the main experiment (Experiment 1). To validate that our experimental results are not driven by a specific choice of parameters, we conducted an additional experiment with a different set of parameters in Appendix E.1 and E.2.

3.3. Experimental Procedure

As presented throughout this Section, we chose an abstract setting in our main experiment for two following reasons. In the literature studying effort provision, two main paradigms exist: one based on *stated effort* and the other based on *real effort* (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 (Bull et al. 1987, Fehr et al. 1993). 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 and to identify empirical anomalies between the theory and the experiment.

Another reason for using an abstract model in our experiment is that data science contests have a wide range of contexts, which may impact the results differently. For example, the level of difficulty of the problem being solved, the different measurements of performance, and the heterogenous ability of contestants can all have a significant impact. To help ensure that our observations on the impact of the two contest design features are applicable to any data science contests in general, and not specific to a particular context in which the contest was held, experimentally testing an abstract model is preferred.

[13] This is one limitation of experimentally testing our abstract model to avoid participants ending up with a negative payoff.

This means that the experiment being conducted is framed in a way that is not tied to any specific industry or domain. This allows for a more controlled and fair interpretation of the results, making it easier to identify the impact of the contest structure and prize structure. One limitation of our approach is that selecting a number may not accurately represent the field environment due to abstraction of the reality.

The whole experiments consist of 352 students at a large research university, with 165 students in our main experiment (Experiment 1) and 187 students in three validation experiments (Appendix E). Each of the four treatments in the main experiment consisted of 39~42 students and each student only participated in one treatment. In each treatment, there are 15 decision rounds. We implemented the experiment using z-Tree software (Fischbacher 2007). The participants earned cash awards ($8 on average) based on their game outcomes (i.e., the sum of payoffs) across 15 rounds. Table 3 maps the terminology used in our base models to the terms used in the experimental instructions for each treatment. Figure 1 shows the decision steps faced by participants. The instruction for the One-High treatment (OH) is in Appendix A.

Table 3. Terminology in the Model and the Experiment

Contest Structure	Model	Experiment	Treatment Abbreviation
All	Contestants	Participants	OH, TH, OL, TL
One-stage	e_i	Decision Number	OH, OL
	ϵ_i	Random Number	OH, OL
	$y_i = e_i + \epsilon_i$	Final Number	OH, OL
Two-stage	$e_{i,1}$	First Decision Number	TH, TL
	$\epsilon_{i,1}$	First Random Number	TH, TL
	$y_{i,1} = e_{i,1} + \epsilon_{i,1}$	First Stage Number	TH, TL
	$e_{i,2}$	Second Decision Number	TH, TL
	$\epsilon_{i,2}$	Second Random Number	TH, TL
	$y_{i,2} = e_{i,2} + \epsilon_{i,2}$	Second Stage Number	TH, TL
	$y_i = y_{i,1} + y_{i,2}$	Final Number	TH, TL
All	$c(e_i)$	Decision Cost	OH, TH, OL, TL
	P_1, P_2, P_3	Award	OH, TH, OL, TL
	π_i	Participant's Point Earning	OH, TH, OL, TL

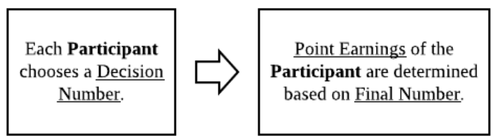
One-stage Contest (OH and OL). At the start of each round, participants needed to enter a Decision Number (e_i) between 1 and 35 into the computer. They were informed that each decision number carries a corresponding Decision Cost ($c(e_i)$). The decision costs were provided in the "Decision Cost Table". Next, the computer would generate a Random Number (ϵ_i) from uniform distribution $unif(-15,15)$ for each participant and calculate the corresponding Final Number ($y_i = e_i + \epsilon_i$). Awards ($P_1 \geq P_2 \geq P_3$) would be given to participants according to their Final Number. Participants with the highest Final Number would be given P_1, and so on. At the end of each round, participants would privately see their rank and their Point Earnings calculated by the Award received minus the Decision Cost incurred.

Two-stage Contest (TH and TL). In the first stage of each round, participants were asked to choose a First Decision Number ($e_{i,1}$) between 1 and 35. There was also a corresponding Decision Cost ($c(e_{i,1})$) for each

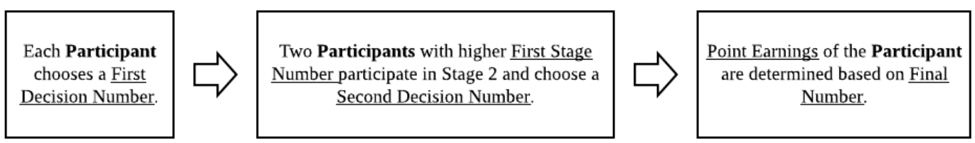
First Decision Number. Next, the computer would generate each participant's First Random Number ($\epsilon_{i,1}$) from uniform distribution $unif(-15,15)$ and calculate the First Stage Number ($y_{i,1} = e_{i,1} + \epsilon_{i,1}$). Then, participants would see whether or not they can participate in the second stage. If their First Stage Number was not the lowest among the three, they could participate in the second stage. In the second stage, two shortlisted participants needed to select a Second Decision Number ($e_{i,2}$) between 1 and 35, which also has a corresponding Decision Cost ($c(e_{i,2})$). The computer would generate a Second Random Number ($\epsilon_{i,2}$) from $unif(-15,15)$ and then calculate the Second Stage Number ($y_{i,2} = e_{i,2} + \epsilon_{i,2}$) as well as the Final Number ($y_i = y_{i,1} + y_{i,2}$). Awards ($P_1 \geq P_2 \geq P_3$) would be given to participants according to their Final Number. At the end of each round, participants would privately see their rank and their Point Earnings.

Figure 1. Steps in Each Round of the Experiment

(a) One-stage Contest (OH, OL)



(b) Two-stage Contest (TH, TL)



3.4. Base Model Prediction with Rational Contestants

Our base model prediction is based on the model equilibrium presented in Section 3.1. If participants are rational, they would behave to maximize their expected payoff and their decision would follow equation (7) for One-High (OH) and One-Low (OL) and equation (13) and equation (12) for the first stage and second stage of Two-High (TH) and Two-Low (TL). Under the same prize structure, the difference in the total effort between *two-stage* contest and *one-stage* contest is given by (see Appendix B for details):

$$\left(e_{i,1}^* + e_{i,2}^*\right) - e_i^* = \frac{[6kv^2 - (P_1 - P_2)](P_1 - P_2)}{144\, k^2 v^3} \in (0, \frac{9}{144} v] \tag{14}$$

Equation (14) suggests that $\frac{e_{i,1}^* + e_{i,2}^*}{e_i^*} \leq 1.083$ and the maximum value is obtained when $P_1 - P_2 = 3kv^2$ and $P_2 = P_3$. In other words, the difference of the total effort between *two-stage* contest and *one-stage* contest cannot exceed 8.3%, which suggests that the efforts from both contests are expected to be rather similar. Moreover, based on equations (7), (12), and (13), *high-spread* prize structure is expected to yield higher effort than *low-spread* prize structure.

Table 4's left part displays the base model predictions with the corresponding parameters adopted for each of the four treatments. To be specific, given that the budget is fixed at 12, the expected total efforts are 18 for One-High and 18.6 for Two-High under *high-spread* prize structure, which are quite similar (3.3% difference). In terms of *low-spread* prize structure, the expected total efforts are 12 for One-Low and 12.7 for Two-Low, again quite similar (5.8% difference) between the two, but they are significantly lower (about 6 points lower for both contest structures) than those of *high-spread* treatments. In other words, the contest structure does not matter much whereas *high-spread* prize structure is surely preferred.

4. Experimental Results

Table 4. Design and Summary of Experimental Results

Contest Structure		Base Model Prediction $\boldsymbol{\beta} = \theta_T = 0$		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	18.0	12.0	**16.5** (11.8) N=585 t=-1.18, p=0.238	**15.0** (9.3) N=630 t=2.55, p=0.011
Two-stage contest	$e_{i,1} + e_{i,2}$	18.6	12.7	**33.3** (16.0) N=420 t=7.23, p<0.001	**24.7** (12.1) N=420 t=7.88, p<0.001
	$e_{i,1}$	6.6	8.7	12.9 (9.5) N=630 t=5.16, p<0.001	12.7 (7.2) N=630 t=4.75, p<0.001
	$e_{i,2}$	12.0	4.0	17.6 (9.8) N=420 t=4.75, p<0.001	10.2 (8.2) N=420 t=5.69, p<0.001
Ratio of Two-stage to One-stage	$\frac{e_{i,1} + e_{i,2}}{e_i}$	**103%**	**106%**	**202%**	**165%**

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 corresponding prediction of the base model. Standard errors are clustered at the participant's level. $\boldsymbol{\beta}$ refers to β_O, β_{T1} and β_{T2}.

The summary of our experimental results for the four treatments is shown in the right-hand panel of Table 4.[14] To ease comparison, we also present the base model predictions for the four treatments in the left-hand panel of Table 4. For *one-stage* treatments, the average effort level is 16.5 for One-High (OH) and 15.0 for One-Low (OL). In *two-stage* treatments, the average total effort is 33.3 for Two-High (TH) and 24.7 for Two-Low (TL). Specifically, in the Two-High treatment, the average of the first-stage effort and that of the second-stage effort are 12.9 and 17.6 respectively. For the Two-Low treatment, the first-stage effort and the second-stage effort are 12.7 and 10.2 on average. Note that the ratio of total effort in *two-stage* contest to that in *one-stage* contest is 202% for *high-spread* prize structure and 165% for *low-spread* prize structure, both much higher than the base model prediction (*High-spread*: 103%, *Low-spread*: 106%).

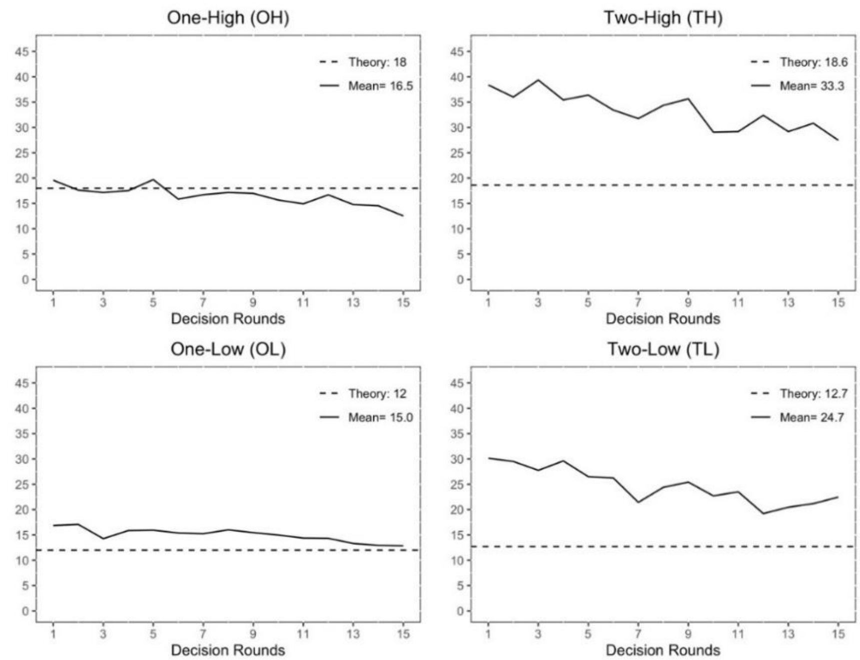
The average of the total effort level in every decision round for each of the four treatments is plotted in Figure 2(a) and, similarly, the average effort level of each stage for each of the two-stage treatments is

[14] Note that, for *two-stage* treatments, since only two thirds of participants are allowed to make a decision in the second stage, the number of effort decisions (N) in the second stage is two-thirds of the number of the first-stage effort decisions.

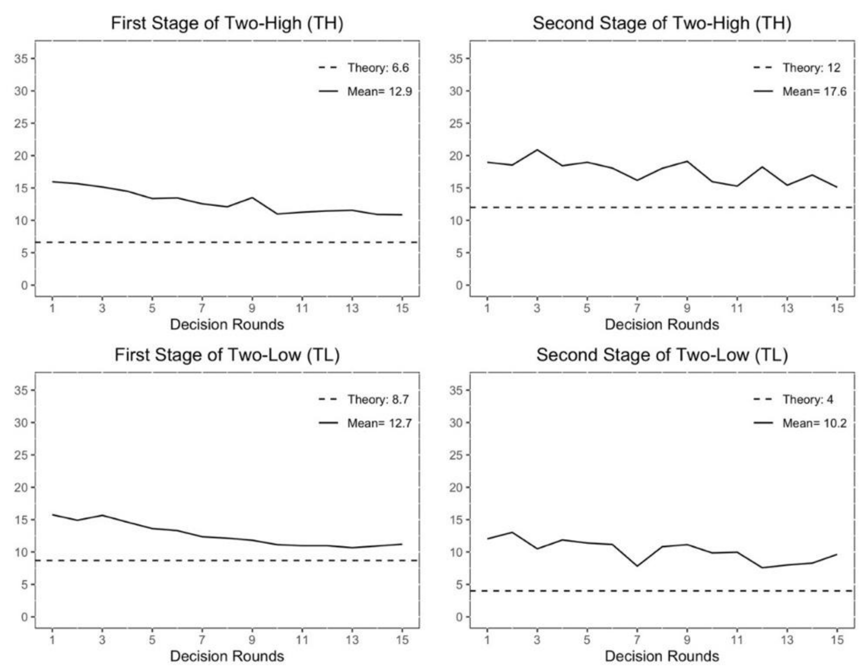
plotted in Figure 2(b). We represent the average effort with a solid line and the base model prediction with a dashed line. We observe that average effort levels (solid line) are always above the base model predictions (dashed line) except for One-High. We now proceed to discuss the formal statistical tests we conducted to examine the impact of contest and prize structures on participants' effort decisions.

Figure 2. Mean Effort Decisions Across 15 Decision Rounds in Main Experiment

(a) Mean Total Effort of Four Treatments



(b) Mean Effort in Each Stage of Two-stage Treatments



4.1. Comparing decisions against the base model.

To compare participants' effort decisions (right panel of Table 4) against the point predictions from the base model (left panel of Table 4), we conducted t-tests. Because participants made multiple decisions across 15 rounds, we clustered the standard errors at the participant's level to account for potential within-subject correlation (Chen et al. 2011, Kim et al. 2019, Chung et al. 2020).[15] The statistical test results are listed in the right panel of Table 4.

The average effort of One-High is 16.5, which appears slightly below, but not significantly different from the base model prediction of 18.0 ($t=-1.18$, $p=0.238$). However, the average effort of One-Low (15.0) is slightly above and significantly different from the prediction of 12.0 ($t=2.55$, $p=0.011$). In terms of *two-stage* treatments, the experimental results of each stage in the two treatments are all significantly greater than the base model predictions. For the first-stage effort, the average effort in Two-High and Two-Low are 12.9 and 12.7, both significantly higher than base model predictions of 6.6 ($t=5.16$, $p<0.001$) and 8.7 ($t=4.75$, $p<0.001$) respectively. The average effort in the second stage for Two-High and Two-Low are 17.6 and 10.2, which are again significantly higher than the predictions of 12.0 ($t=4.75$, $p<0.001$) and 4.0 ($t=5.69$, $p<0.001$). As a result, the total effort (which combines two effort levels from both the first stage and the second stage) for *two-stage* treatments for two different prize structures are at least 12 points higher (79% more) than the base model predictions (Two-High: $t=7.23$, $p<0.001$; Two-Low: $t=7.88$, $p<0.001$). Figure 2 also displays similar patterns. Across all 15 decision rounds, the average effort in each round (solid line) is located above base model prediction (dashed line) in all the treatments except for One-High.[16]

In sum, comparing against the base model predictions, contestants significantly boost their effort in both stages of the *two-stage* treatments whereas, in the *one-stage* treatments, contestants' effort level is close to (or slightly higher than) the base model prediction. When we compare effort levels in the first half rounds (Rounds 1-8) and the second half rounds (Rounds 9-15) separately against the above predictions for each treatment, the results remain largely the same (see Appendix C). These results suggest that contestants are also motivated by nonpecuniary preferences when making decisions.

4.2. Comparing decisions between one-stage and two-stage contest structure.

This section aims to examine the following question: Given a fixed budget of prize pool, which contest structure (*one-stage* vs. *two-stage*) would let contestants exert more effort? Note that the base model suggests that the contest structure does not matter much. That is, under the same prize structure, *one-stage*

[15] We did this for all the statistical tests reported in this paper. We also performed Wilcoxon Rank-Sum test and the results are consistent.

[16] Figure 2 suggests presence of learning behavior in in the lab experiments. However, it is unlikely that contestants would significantly lower their efforts by participating in a large number of competitions over a short period of time in a real data science competition platform. For example, Kaggle users participate in 2.8 data science competitions on average, based on the metadata provided by Kaggle (see https://www.kaggle.com/code/jtrotman/meta-kaggle-count-user-activities. Retrieved on 30th Jan, 2023). We thank the SE and the anonymous reviewer for this insightful observation.

contest and *two-stage* contest should yield similar effort levels (less than 8.3% difference) in theory. To answer the above question, we conduct pairwise t-tests between the two contest structures.

We first compare the average total effort levels between two different contest structures under the same prize structure. Table 5 summarizes the t-test results. *Two-stage* treatments significantly outperform *one-stage* counterparts under both *high-spread* prize structure ($\Delta\bar{e} = 16.8$, t=7.04, p<0.001) and *low-spread* prize structure ($\Delta\bar{e} = 9.7$, t=5.08, p<0.001), which is significantly larger than the base model predictions. In the experiment, the total effort for *two-stage* treatments is 165% (*low-spread*) to 202% (*high-spread*) of the effort in *one-stage* treatments although the base model suggests total effort in *two-stage* contest should be only 3% (*high-spread*) to 6% (*low-spread*) higher. These suggest that, given a fixed budget, the contest organizer should choose to implement *two-stage* contest where they let all contestants participate in the first stage and make the shortlisted contestants continue to solve the same problem.

Table 5. Comparison between Different Contest Structures

Prize Structure	Experiment Result: Average Total Effort			t-statistics	*p*-values
	Contest Structure: One-stage Contest	Contest Structure: Two-stage Contest	Ratio of Two-stage to One-stage		
High-spread	16.5	33.3	**202%**	-7.04	<0.001
Low-spread	15.0	24.7	**165%**	-5.08	<0.001

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

Notice that, while our results are directionally in line with the base model predictions, the magnitudes of differences are much greater than expected. This observation is quite surprising given that many papers in experimental OM show that the differences observed from experimental results are often smaller than theory predictions. For example, Davis et al. (2014) found that experimental results of push and pull contracts qualitatively agree with the theory, but the actual level of profit difference is smaller. Similarly, Davis (2015) experimentally found that the benefit of the coordinating contracts over the wholesale price contract is less than the standard theory predicts. These indicate that behavioral components play an important role in participants' decision making, especially in the *two-stage* treatments.[17]

4.3. Comparing decisions between high-spread and low-spread prize structure.

Next, we conduct t-tests between different prize structures while keeping the same contest structure to answer the following question: Which prize structure would yield a higher effort level? The base model suggests that the *high-spread* prize structure is better. Kalra and Shi (2001) demonstrated that if shocks follow uniform distribution, the *winner-takes-all* prize structure would yield the best performance. In our context, *high-spread* prize structure best mimics the *winner-takes-all* prize structure because the 2nd and the 3rd rank can only win the award $P_2 = P_3 = 2$ and most money goes to the 1st rank.

[17] We thank the SE and the anonymous reviewer for this sharp observation.

The left-panel of Table 6 shows the average effort for each of the four treatments. The right-panel of Table 6 displays the results of t-tests between two different prize structures under the same contest structure. For *one-stage*, the prize structure does not influence contestants' decisions on effort ($\Delta\bar{e} = 1.5$, t=0.87, *p*=0.384), which is in line with Lim et al. (2009)'s findings. In terms of *two-stage*, prize structure does not influence the first-stage decisions whereas it influences the second-stage decisions. The second-stage effort of *high-spread* is significantly higher than that of *low-spread* prize structure ($\Delta\bar{e} = 7.4$, t=4.63, *p*<0.001). In consequence, the total effort of *high-spread* is also significantly higher than that of *low-spread* prize structure ($\Delta\bar{e} = 8.6$, t=3.39, *p*=0.001). As Kalra and Shi (2001) showed that the *winner-take-all* prize structure is better in theory, our experimental results show that the *winner-take-all* prize structure indeed works well in *two-stage* contests. Our experimental results clearly show that the number of stages should be considered first in determining which prize structure is preferred.[18]

Table 6. Comparison between Different Prize Structures

Contest Structure	Stage	Experiment Result: Average Effort		t-statistics	*p*-values
		Prize structure: High-spread	Prize Structure: Low-spread		
One-stage Contest	e_i	16.5	15.0	0.87	0.384
Two-stage Contest	$e_{i,1} + e_{i,2}$	33.3	24.7	3.39	0.001
	$e_{i,1}$	12.9	12.7	0.14	0.887
	$e_{i,2}$	17.6	10.2	4.63	<0.001

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

4.4. Comparing decisions between first stage and second stage in two-stage contest.

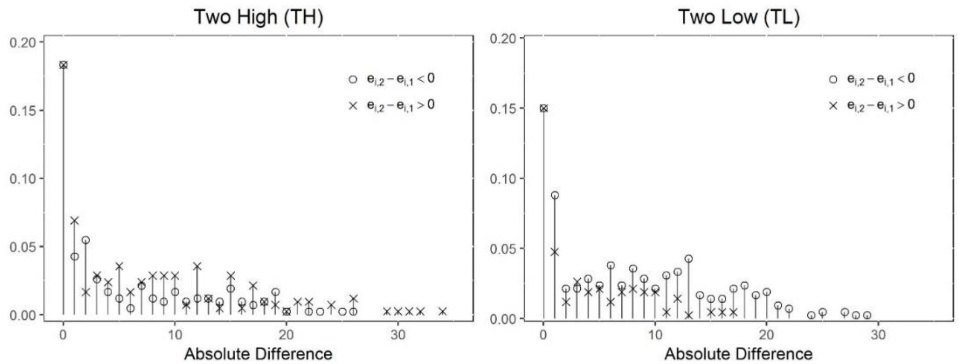
In this section, we provide some suggesting evidence that contestants tend to stick with their previous decisions. We plot the histogram of the absolute difference between the second-stage effort and the first-stage effort $|e_{i,2} - e_{i,1}|$ in Figure 3(a) and superimpose dots showing whether $e_{i,2} - e_{i,1}$ is positive (cross dots) or negative (circle dots). As is shown in Figure 3(a), sticking with previous decision ($e_{i,2} - e_{i,1} = 0$) is most frequently observed in both Two-High treatment and Two-Low treatment. Furthermore, based on the absolute difference of $e_{i,2} - e_{i,1}$, we categorize contestants into several categories: 0~5 ($|e_{i,2} - e_{i,1}| \leq 5$), 5~10 ($5 < |e_{i,2} - e_{i,1}| \leq 10$), 10~15 ($10 < |e_{i,2} - e_{i,1}| \leq 15$), 15~20 ($15 < |e_{i,2} - e_{i,1}| \leq 20$), 20~25 ($20 < |e_{i,2} - e_{i,1}| \leq 25$), 25~30 ($25 < |e_{i,2} - e_{i,1}| \leq 30$), >30 ($|e_{i,2} - e_{i,1}| > 30$) and calculate the proportion that falls into each category at each decision round in Figure 3(b). Category 0~5 represents the contestants whose second-stage effort does not deviate too much from their first-stage effort. Figure 3(b) shows that the proportion of category 0~5 is around 50% across all decision rounds (Two-High: 51%, Two-

[18] Given that the prize structure is only slightly different across treatments, the difference in effort provision between *high-spread* and *low-spread* prize structure of *two-stage* contest is significant. Our *two-stage* results may be even larger under winner-takes-all design where the last prize $P_3 = 0$. Similarly, this indicates that the difference in *one-stage* treatments might be understated. We thank the SE and the anonymous reviewer for this insightful comment.

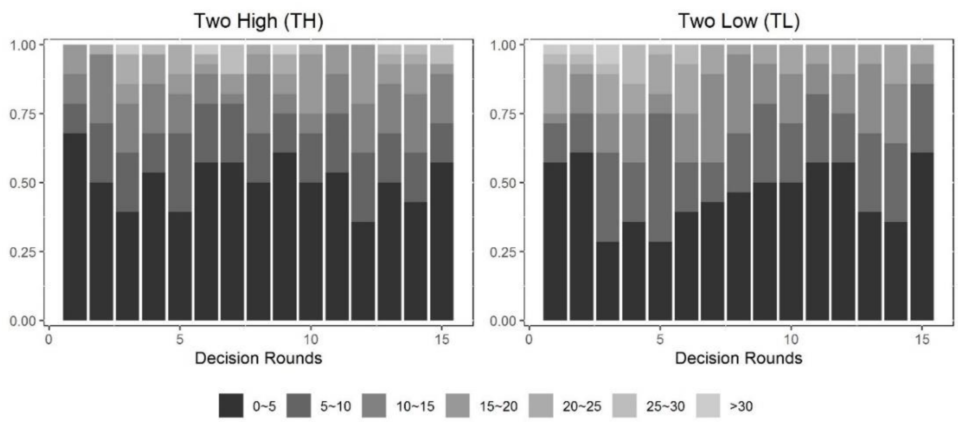
Low: 46% in total), suggesting that most contestants avoid deviating much from their previous decisions.

Figure 3. Distribution of $|e_{i,2} - e_{i,1}|$

(a) Histogram of $|e_{i,2} - e_{i,1}|$



(b) Proportion of $|e_{i,2} - e_{i,1}|$ across Decision Rounds



4.5. Summary of empirical regularities.

Overall, we show the following empirical regularities: *Two-stage* contests achieve higher effort levels from contestants than *one-stage* contests do. *High-spread* prize structure is better under *two-stage* contests whereas prize structure does not matter for *one-stage* contests. More specifically, as discussed earlier, a detailed examination of experimental results reveals that contestants boost their effort in both of the first stage (Two-High: $\Delta\bar{e} = 6.3$, t=5.16, p<0.001; Two-Low: $\Delta\bar{e} = 4$, t=4.75, p<0.001) and the second stage (Two-High: $\Delta\bar{e} = 5.6$, t=4.75, p<0.001; Two-Low: $\Delta\bar{e} = 6.2$, t=5.69, p<0.001) of the *two-stage* contest whereas the effort levels do not deviate much from the base model predictions in the *one-stage* contest. Moreover, the total effort of *two-stage* contest is almost twice of the total effort of *one-stage* contest. These empirical regularities cannot be explained by predictions from our base model. In the next section, we propose a behavioral model with two non-pecuniary factors that help explain the contestants' behaviors.

5. Behavioral Model

While standard theory assumes that people behave rationally to maximize their own monetary payoff, the recent developments in behavioral OM suggest that many choices are driven by both monetary payoff and behavioral factors (Mullainathan and Thaler 2000, Camerer 2011, Davis et al. 2014, Davis 2015, Becker-Peth et al. 2020). Our behavioral model is also motivated by such possibility that the contest structure may psychologically influence contestants' decisions. The previous literature on tournaments has shown that aversion to being responsible for teams' loss can yield higher effort (Chen and Lim 2013), and contestants often care about their outcome relative to other contestants (Lim 2010). Relatedly, Roels and Su (2014) proposed the concept of "behind aversion" to characterize situations where people may face disutility from underperforming relative to others. In our context, it could be that, in *two-stage* design, being eliminated from further competition after the first stage can be seen as a clear sign of failure and falling behind. If they cannot participate in the second stage, they will feel that they are underperforming relative to other contestants who are allowed to participate in the second stage. This psychological aversion to being eliminated, also conceptualized as "behind aversion" (Roels and Su 2014) will motivate contestants to exert more effort in order to enter the second stage. Therefore, we include the first parameter of interest β to capture a contestant's disutility from losing the competition – i.e., whether the contestant is averse to being left behind.

In addition, in multi-stage contests, contestants who continue to participate in the next stage may continue exerting significant effort in the second stage. Previous literature suggests that people have a tendency to stick with their previous decisions (Samuelson and Zeckhauser 1988, Fernandez and Rodrik 1991, Masatlioglu and Ok 2005). The concept of this phenomenon has been used to explain many irrational behaviors of firms as well as individuals. For example, organizations sometimes resist implementing large-scale information technologies or replacing an incumbent system because of status quo (Kim and Kankanhalli 2009, Polites and Karahanna 2012). With regard to healthcare choices, mothers usually stick to maintaining the traditional infant healthcare practice (Venkatesh et al. 2016). Long et al. (2020) show that status quo explains delaying project abandonment. In our context, contestants are likely to maintain the status quo and, as a result, they are averse to deviating from their previous decisions (i.e., continuously over-exerting effort). Thus, the second parameter θ in our behavioral model measures the psychological disutility from deviating from their previous decisions. This parameter exists only in *two-stage* contests. Our behavioral model nests the base model if all the behavioral parameters are zero. We specify the behavioral model for *one-stage* contests and *two-stage* contests in greater detail in Sections 5.1 and 5.2 respectively.

5.1. Contestant's Utility in One-stage Contest

Here, we start by presenting the behavioral model for *one-stage* contest first. A critical question when modeling a contestant's psychological utility loss from being left behind is how a reference point is formed by the contestant to make a comparison with others (Tversky and Kahneman 1979, Loewenstein et al. 1989, Roels and Su 2014). The most natural way would be that the contestant evaluates his prize against the average value of all available prizes (Lim 2010, Roels and Su 2014). Thus, by incorporating that the contestant may experience psychological loss when his prize is lower than the average prize P_{avg}, we propose the utility function for contestant i at round t of *one-stage* contest as:

$$U_{i,t}\left(e_{i,t}|\beta_O\right) = \sum_{r\in\{1,2,3\}} Prob\left(P_{i,t} = P_r\right) \times \left\{P_r - I\left(P_{avg} > P_r\right) \times \beta_O \times \left(P_{avg} - P_r\right)\right\} - c\left(e_{i,t}\right) \quad (15)$$

The first component of equation (15) is a sum of expected returns from the contest being ranked at the first, the second, or the third while the last component is a decision cost ($c\left(e_{i,t}\right)$). In specific, $Prob\left(P_{i,t} = P_r\right)$ refers to the probability of the contestant i being ranked at r^{th} position in round t as discussed in Section 3.1. The utility of winning the r^{th} prize equals to the monetary prize P_r minus the disutility from being left behind. Notice that the contestant experiences this disutility (($\beta_O \times \left(P_{avg} - P_r\right)$) only when the awarded prize is lower than the average (i.e., $P_{avg} > P_r$). $\beta_O > 0$ captures the existence of disutility from underperforming relative to others in *one-stage* contest. The magnitude of the disutility depends on the difference between the average prize P_{avg} and the awarded prize P_r. Basically, the lower the prize the contestant receives, the higher the disutility the contestant experiences.

By solving the first order condition of the utility function in equation (15), we have the following equilibrium:

$$e_i^* = \frac{P_1 - P_3 + \beta_O\left(P_{avg} - P_3\right)}{4kv} \quad (16)$$

If $\beta_O = 0$, equation (16) nests the base model (i.e., equation (7)) as a special case where the contestant is rational and only cares about the monetary components.

5.2. Contestant's Utility in Two-stage Contest

In this section, we provide the behavioral model for *two-stage* contest. In addition to the behind aversion parameter β introduced in our model for *one-stage* contest, we also include a parameter θ which captures contestants' tendency to stick with previous decisions in the second stage of the *two-stage* contest. Here, we first specify the contestant's utility function in the second stage ($k = 2$) of *two-stage* contest:

$$U_{i,2,t}\left(e_{i,2,t}\middle|\theta_T, e_{i,1,t}\right)$$
$$= \sum_{r\in\{1,2\}} Prob\left(P_{i,t} = P_r\right) \times \left\{P_r - I\left(P'_{avg} > P_r\right) \times \beta_{T2} \times \left(P'_{avg} - P_r\right)\right\} - \theta_T\left(e_{i,2,t} - e_{i,1,t}\right)^2 - c\left(e_{i,2,t}\right) \quad (17)$$

In the second stage of two-stage contest, only two shortlisted contestants can participate. The first component measures the utility from receiving the first rank and the second rank. $Prob\left(P_{i,t} = P_r\right)$ refers to the probability of ranking at r^{th} position in round t and the utility equals the monetary prize P_r minus the additional disutility ($\beta_{T2} \times \left(P'_{avg} - P_r\right)$) only when the awarded prize is lower than the average (i.e., $P'_{avg} >$

P_r). Here P'_{avg} is the average prize of the second stage. $\beta_{T2} > 0$ captures the existence of disutility from underperforming relative to others in the second stage of *two-stage* contest. $-\theta_T\left(e_{i,2,t} - e_{i,1,t}\right)^2$ captures the contestants' tendency to stick with previous decisions.[19] If the contestant's second-stage effort $e_{i,2,t}$ deviates from his first-stage effort $e_{i,1,t}$, he would incur a negative utility ($\theta_T > 0$). The higher the difference between the second-stage effort and the first-stage effort, the larger the disutility. $\theta_T > 0$, thus, captures the degree of the aversion to deviating from previous decisions. The last component $c\left(e_{i,2,t}\right)$ is the second-stage decision cost.

Now, we show the contestant's utility function in the first stage of the *two-stage* contest as follows:

$$U_{i,1}\left(e_{i,1,t}\middle|\beta_T\right) = Prob\left(P_{i,t} = P_3\right) \times \left\{P_3 - \beta_{T1}\left(P_{avg} - P_3\right)\right\} \\ +\left\{1 - Prob\left(P_{i,t} = P_3\right)\right\} \times \left\{\frac{P_1 + P_2}{2} - c\left(e^*_{i,2}\right)\right\} - c\left(e_{i,1,t}\right) \tag{18}$$

$Prob\left(P_{i,t} = P_3\right)$ is the probability of being eliminated in the first stage. In that case, the contestant receives P_3 and $\beta_{T1} > 0$ captures the utility loss from not being shortlisted (i.e., losing). Here, the reference point is the average prize $P_{avg} = 4$. Next, $1 - Prob\left(P_{i,t} = P_3\right)$ refers to the chance of being shortlisted. $\frac{P_1+P_2}{2} - c\left(e^*_{i,2}\right)$ is the expected payoff ($E\pi_{i,2}$) that the contestant would get if he continues to participate in the second stage where $e^*_{i,2}$ is the contestant's second stage equilibrium. Lastly, $c\left(e_{i,1,t}\right)$ is the first-stage cost.

Solving the first order condition of equation (17) yields the second-stage equilibrium as:

$$e^*_{i,2} = \frac{\left(P_1 - P_2 + 6\theta_T v e_{i,1} + \beta_{T2}(P'_{avg} - P_2)\right)}{6(k + \theta_T)v} \tag{19}$$

Plugging equation (19) into equation (18), the first order condition of equation (18) is as follows:

$$Ae^2_{i,1} + Be_{i,1} + C = 0 \tag{20}$$

where A, B, and C are:

$$A = -\frac{\theta_T^2}{4(k + \theta_T)^2 v}, B = -1 - \frac{(P_1 - P_2)\theta_T}{12(k + \theta_T)^2 v^2} - \frac{\left(\beta_{T2}\left(P'_{avg} - P_2\right)\right)\theta_T}{12(k + \theta_T)^2 v^2},$$

$$C = -\frac{\left(\beta_{T2}\left(P'_{avg} - P_2\right)\right)^2}{144(k + \theta_T)^2 v^3} - \frac{\beta_{T2}\left(P'_{avg} - P_2\right)(P_1 - P_2)}{72(k + \theta_T)^2 v^3} - \frac{(P_1 - P_2)^2}{144(k + \theta_T)^2 v^3} + \frac{\beta_{T1}\left(P_{avg} - P_3\right)}{4kv} \\ + \frac{P_1 + P_2 - 2P_3}{8kv}$$

The first-stage equilibrium will be the solution of equation (20).[20] Note that the behavioral

[19] The reason why we use squared deviation is to make the closed form solution of the optimal second-stage effort $e^*_{i,2}$ be a function of $e_{i,1}$ so that we can link it with our behavioral theory (i.e., a function of previous first-stage effort).

[20] Though it is straightforward to infer the closed-form solution of the first-stage equilibrium, the functional form itself is rather complex to present. Thus, we skip reporting the closed-form solution here in the manuscript.

equilibrium nests the base model as a special case if $\beta_{T1} = \beta_{T2} = 0$ and $\theta_T = 0$.

5.3. Estimating the Behavioral Model

We utilize the effort level decisions (e_i) made by the contestants from the four treatments in our experiment to estimate the behavioral parameters in equations (16), (19) and the solutions to equation (20) via maximum likelihood method. Specifically, we assume that for *one-stage* contest $e_{i,t,TR} \sim N(e^*_{TR}, \sigma^2_{TR})$ and for *two-stage* contest $e_{i,k,t,TR} \sim N(e^*_{TR,k}, \sigma^2_{TR,k})$, where i represents contestant i, t is the decision round, TR indicates the four treatments, k denotes stage k, e^*_{TR} is the equilibrium effort of *one-stage* contest in treatment TR, σ^2_{TR} is the variance of *one-stage* contest in treatment TR, $e^*_{TR,k}$ is the equilibrium effort of stage k in *two-stage* contest, $\sigma^2_{TR,k}$ is the variance of stage k in *two-stage* contest, and $f(.)$ is PDF of the Normal distribution. We have 4 treatments in total – One-High (OH) and One-Low (OL) for *one-stage* contest and Two-High (TH) and Two-Low (TL) for *two-stage* contest. In each treatment, there are N_{TR} contestants who participate in 15 decision rounds. Therefore, the joint likelihood function is given by:

$$L(\beta_O, \beta_{T1}, \beta_{T2}, \theta_T, \sigma_{TR}, \sigma_{TR,k}) = \prod_{TR}^{\{OH,OL\}} \prod_{i=1}^{N_{TR}} \prod_{t=1}^{15} f\left(N(e^*_{TR}, \sigma^2_{TR})\right) \prod_{TR}^{\{TH,TL\}} \prod_{i=1}^{N_{TR}} \prod_{t=1}^{15} \prod_{k=1}^{2} f\left(N(e^*_{TR,k}, \sigma^2_{TR,k})\right) \quad (21)$$

5.4. Estimation results

The parameter estimates of the behavioral model and nested models are shown in Table 7. As we will discuss, the parameter estimates confirm our behavioral model. We first examine the parameter estimates of our proposed behavioral model (see "Full model" shown in column (1) of Table 7). For *one-stage* contest, $\beta_O = 0.227$ is significantly different from zero (p<0.001), implying that contestants in *one-stage* contest experience slight "behind aversion". This is not surprising because the experimental results of *one-stage* contest are slightly higher than the base model predictions.

For *two-stage* contest, $\beta_{T1} = 1.095$ is positive and statistically different from zero (p<0.001), showing that the behavior of contestants is consistent with our behavioral model that incorporates psychological aversion to being left behind at the elimination stage. That is, contestants exhibit disutility when they are not shortlisted for the second stage. Therefore, they tend to exert higher effort in the first stage than the base model prediction. Note that the magnitude of β_{T1} is much larger than β_O (W=158.85, p<0.001), suggesting that contestants experience much more disutility from ranking 3rd in *two-stage* contest than ranking 3rd in *one-stage* contest. The reason is that being eliminated in *two-stage* contest is a more salient sign of losing compared with ranking 3rd in *one-stage* contest. We also note that, in the second stage of *two-stage* contest, contestants still exhibit aversion to falling behind ($\beta_{T2} = 1.727$, p<0.001). In addition, in the second stage, $\theta_T = 0.006$ is positive and statistically significant (p<0.001), which suggests that contestants avoid deviating from their first-stage decisions. They tend to stick with their previous decisions. Behind aversion together with a tendency to stick with previous decisions, both motivate the contestant in the second stage to exert much higher effort than what the base model predicts.

Table 7. Estimates of the Behavioral Model and Comparison with Nested Models

	Parameter	(1) Full model	(2) Nested model 1 ($\beta_{T1} = \beta_{T2}$)	(2) Nested model 2 ($\beta_O = \beta_{T1} = \beta_{T2}$)	(3) Nested model 3 ($\boldsymbol{\beta} = 0$)	(4) Nested model 4 ($\theta_T = 0$)	(5) Nested model 5 ($\boldsymbol{\beta} = \theta_T = 0$)
One-stage contest	β_O	0.227***	0.227***	0.758***		0.227***	
		(0.052)	(0.052)	(0.034)		(0.052)	
	σ_{OH}	12.088***	12.088***	13.212***	11.838***	12.088***	11.838***
		(0.361)	(0.361)	(0.398)	(0.346)	(0.361)	(0.346)
	σ_{OL}	9.438***	9.438***	9.427***	9.766***	9.438***	9.766***
		(0.271)	(0.271)	(0.268)	(0.275)	(0.271)	(0.275)
Two-stage contest	β_{T1}	1.095***	1.113***			0.968***	
		(0.045)	(0.045)			(0.045)	
	β_{T2}	1.727***				0.849***	
		(0.224)				(0.067)	
	θ_T	0.006***	0.005***	0.005***	0.006*		
		(0.001)	(0.001)	(0.001)	(0.003)		
	$\sigma_{TH,1}$	9.658***	9.569***	9.808***	10.946***	9.957***	11.399***
		(0.277)	(0.271)	(0.280)	(0.315)	(0.293)	(0.321)
	$\sigma_{TH,2}$	9.933***	10.128***	10.636***	12.685***	9.843***	11.320***
		(0.348)	(0.358)	(0.386)	(0.504)	(0.340)	(0.391)
	$\sigma_{TL,1}$	7.220***	7.273***	7.172***	8.369***	7.336***	8.172***
		(0.206)	(0.209)	(0.202)	(0.240)	(0.214)	(0.230)
	$\sigma_{TL,2}$	8.216***	8.189***	8.302***	9.074***	9.348***	10.276***
		(0.285)	(0.283)	(0.292)	(0.357)	(0.329)	(0.355)
Log likelihood		-12098.9	-12104.4	-12188.8	-12424.6	-12178.5	-12439.5
LR test			11.02***	179.84***	651.26***	159.14***	681.12***

Notes: $\boldsymbol{\beta}$ refers to β_O, β_{T1} and β_{T2}. For each nested model, LR test was conducted against the full model. Standard errors are shown in parentheses. **p*<0.05; ***p*<0.01; ****p*<0.001.

The other columns of Table 7 display the fit of various nested models. The results of the likelihood-ratio (LR) tests indicate that the full model with all behavioral parameters produces the best fit. Nested Model 1 set the behind aversion parameter β to be the same in *two-stage* contest ($\beta_{T1} = \beta_{T2}$), which is rejected (*p*<0.001). Nested model 2 sets parameters β to be the same across the two contest structures ($\beta_O = \beta_{T1} = \beta_{T2}$). The model is rejected (*p*<0.001), suggesting that the psychological aversion to falling behind the competition are different for different contest structures. Nested model 3 imposes the restriction of not having behind aversion parameters ($\beta_O = \beta_{T1} = \beta_{T2} = 0$). This model is rejected (*p*<0.001), indicating that incorporating behind aversion factors in the model is necessary. Nested model 4 does not include any bias towards previous decisions ($\theta_T = 0$) and again is rejected (*p*<0.001), suggesting that this psychological factor is also essential in the model. Lastly, nested model 5 is our base model without any behavioral parameters ($\beta_O = \beta_T = \theta_T = 0$). The LR test rejects this nested model as well (*p*<0.001), clearly showing that the full model performs significantly better.

In sum, our behavioral model confirms the following findings: Contestants in *one-stage* contest exhibit slight behind aversion. For *two-stage* contest where the first stage is the elimination stage, contestants are much more averse to being eliminated and the magnitude of behind aversion in *two-stage* contest is more than 4.8 times larger than that in *one-stage* contest. Moreover, in the second stage of *two-stage* contest, shortlisted contestants display continuous exertion of efforts in addition to behind aversion. They tend to stick with their first-stage decisions and continue to exert higher efforts than the base model predictions in the second stage. Overall, our behavioral model incorporating psychological aversion to falling behind and continuous exertion of efforts can better explain the empirical regularities in our experiments.

5.5. Alternative Explanations

Although our behavioral model in Section 5.1 and Section 5.2 fits the experiment results well, there could be other alternative behavioral explanations for our results. We briefly discuss some of those potential alternatives here and share their similarity as well as difference from our context. First, instead of "behind aversion" utilized in our behavioral model, "last-place aversion" could be another potential behavioral factor in explaining our results. Contestants might be averse to being ranked in the "last place" (Kuziemko et al. 2014, Buell 2021). Kuziemko et al. (2014) demonstrate that individuals are averse to being in the "last place" such that if customers are waiting in the last place of a queue, they are more likely to switch or even abandon the queue (Buell 2021). Although the concept of "last-place aversion" seems similar to "behind aversion" in that people are averse to being in the last rank, "last-place aversion" occurs when people first realize their rank (i.e., last place) and then subsequently choose an action (Bull et al. 1987, Kuziemko et al. 2014). However, in our context, participants first choose an action before realizing their rank.

Second, one may consider that the continuous exertion of effort might be due to "anchoring bias" which explains phenomena that people's judgments and decisions are influenced by an initially presented value or reference point (Tversky and Kahneman 1974). For example, initially marketed price is often taken as an anchor by buyers. In this case, the buyers often perceive negotiated price as fair when the initial price is high, even if the negotiated price is more than the product's true market worth (Northcraft and Neale 1987). Although our second-stage effort is affected by the first-stage effort, the first-stage effort is a decision made by contestants themselves while anchors are often exogenous information given to the subjects.

Third, another alternative explanation for higher efforts in *two-stage* contest would be mental accounting. Mental accounting explains the tendency that people psychologically compartmentalize their spending of money into multiple mental accounts based on subjective criteria (i.e., purpose of spending) (Thaler 1999). Similarly, participants in our *two-stage* contest may just treat the two stages separately as two independent games. To test this, we compared decisions between *one-stage* and the separate stages of the *two-stage* treatment. If contestants treat the two stages as two separate *one-stage* contests, we would observe that they are not significantly different from each other. We report the statistical tests of this in Table D of Appendix D, which are not in line with this conjecture.

Lastly, another similar phenomenon to our findings is the "winner's curse" in auction design, where the winning bidder often overbids the product (Thaler 1988). Although participants overinvest both in our *two-stage* experiment and in auction design, we do not find strong evidence that contestants overinvest and diverge from the base model predictions in the case of *one-stage* contest.

As we briefly discussed, while several potential alternative explanations may exist, our behavioral model including behind aversion and continuous exertion of efforts is suitable in explaining behaviors in data science competitions largely due to the following observations: (1) contestants first spend an effort and then realize the rank; (2) the first-stage effort is not exogenous information; (3) contestants' behaviors in the two stages of *two-stage* contest are different; (4) contestants are less likely to overinvest in one-stage contest. Extending this point forward, further studies systematically investigating other potential behavioral factors behind contestants' behaviors would be helpful in enriching our understanding.

6. Validation Experiments

To further demonstrate the robustness of our results, we conduct three additional experiments as follows. In Experiment 2, we incorporate the risk preference of contestants and use a completely new set of parameters to test whether our experimental results are robust and whether our findings are not specific to a certain choice of parameters. The detailed experimental design and results are reported in Appendix E.1. We show that the risk-preference model cannot explain the observed behaviors. Moreover, our findings are robust to different model parameters. In Experiment 3, we vary the number of participants in a contest and confirm that our findings 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. 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 (both in Experiment 2 and 3). If our proposed behavioral model is robust, it should be able to predict contestants' behaviors under other experimental parameters as well. We confirm and report these details in Appendix E.2. Lastly, as mentioned earlier, one downside of our experimental approach could be from a gap between an abstract model and the field environment where real efforts are invested. In Experiment 4, we utilized the real-effort task (Gill and Prowse 2012) and also varied the prize structure by allowing $P_3 = 0$ to better mimic real-world scenarios. The experimental results are in line with our main experiment. Details can be found in Appendix E.3.

7. Conclusion

This paper examines a critical question that every contest organizer faces when they design an open competition especially for data science competitions (e.g., machine learning or big data contests on Kaggle): What is a better contest structure to motivate contestants to exert more effort and, as a result, to achieve a better solution? More specifically, should the contest have multiple stages and which prize structure should be adopted? To answer these questions, we utilize an incentive-aligned experiment which allows us to observe the effort provision from contestants whereas we cannot do so in the field. Our experiment manipulates two types of contest design: (1) the number of stages (*contest structure*) and (2) the distribution of prizes (*prize structure*). Contrary to the base model prediction, our experimental results show that contestants significantly boosted their effort in the *two-stage* contests. Furthermore, we found that contestants provide higher level of effort when the winner takes most of the prizes in the *two-stage* contest. To explain the empirical anomalies observed in our experiment, we developed a behavioral model that captures contestants' psychological aversion to falling behind and tendency to stick with previous decisions. In the first stage of *two-stage* contest, contestants are averse to being eliminated and, thus, they over-exert their effort. In the second stage, contestants stick with their previous decision and keep investing higher effort in the second stage.

Our findings provide several policy implications by demonstrating that it is crucial for contest holders to be aware that contestants' decisions can be influenced by non-pecuniary factors. First of all, an immediate policy implication is that contest organizers should adopt a multi-stage contest by providing partial data in the first stage (and full data in the second stage) whenever possible, which can motivate contestants to invest more effort. Next, having multiple stages in a contest significantly boosts the effort provision by contestants because no one wants to be perceived as a "loser" by being eliminated early. Utilizing this, the contest organizers may make "losing" in the elimination stage more salient (e.g., announcing the contest results publicly), which may lead to even higher effort provision from contestants by strengthening contestants' psychological aversion to falling behind. Furthermore, if the contestants already invested a considerable amount of effort in their elimination stage, they would continue to invest high effort in the following stage. This suggests that the first-stage problem in the contest should be considered challenging from the perspective of contestants.

Because our paper is one of the few studies that have considered the contest design of data science competitions, it is not without limitations and we discuss several ways to extend our work. First, in a certain industry, firms sometimes adopt more than two stages (e.g., *three-stage* or even more stages) in competition. It is therefore worthwhile to empirically examine how contestants' behavior changes as the number of elimination stages increases. Will it motivate the contestants more or demotivate them? Will there be any optimal number of stages? Second, this paper hints that contestants' behaviors are influenced by two factors: psychological aversion to falling behind and a tendency to adhere to previous decisions. Extending this to an individual level, the underlying mechanism of contestants with the increased effort caused by stage split and that of those without could be further investigated. Third, in the real world, contestants' abilities are not homogenous. Our model has a room to be further extended to allow for heterogeneous contestants. That is, what will happen if contestants know that there is a superstar contestant participating in a competition? How would the distribution of ability among contestants have an impact on the effort provision? Fourth, in some open-source competitions, there is a case where the solution of the first-stage winner is disclosed to the public before starting the second stage. It would be interesting to study how contestants will strategically respond in this case. Lastly, as previously discussed, lack of real-world data science context (due to the abstraction in our experiments) could question whether our findings can be applied in the real-life data science competitions. Thus, further investigation and verification of our results in the field would be an important next step.

Acknowledgements

Both authors contributed equally. The authors thank the DE, the SE, and the anonymous reviewers for their constructive feedback. The authors thank participants at SCECR 2020 and 2020 INFORMS Annual Meeting for helpful comments.

References

Ales L., S. H. Cho, E. Körpeoğlu. 2021. Innovation Tournaments with Multiple Contributors. *Production and Operations Management*, **30** (6), 1772-1784.

Alpern S., J. Howard. 2017. Winner-Take-All Games: The Strategic Optimisation of Rank. *Operations Research*, **65** (5), 1165-1176.

Altmann S., A. Falk, M. Wibral. 2012. Promotions and Incentives: The Case of Multistage Elimination Tournaments. *Journal of Labor Economics*, **30** (1), 149-174.

Amaldoss W., A. Rapoport. 2009. Excessive Expenditure in Two-Stage Contests: Theory and Experimental Evidence. *Game Theory: Strategies, Equilibria, and Theorems.* Nova Science Pub Inc, Hauppauge, NY.

Becker-Peth M., K. Hoberg, M. Protopappa-Sieke. 2020. Multiperiod Inventory Management with Budget Cycles: Rational and Behavioral Decision-Making. *Production and Operations Management*, **29** (3), 643-663.

Bimpikis K., S. Ehsani, M. Mostagir. 2019. Designing Dynamic Contests. *Operations Research*, **67** (2), 339-356.

Bolton G. E., E. Katok. 2008. Learning by Doing in the Newsvendor Problem: A Laboratory Investigation of the Role of Experience and Feedback. *Manufacturing & Service Operations Management*, **10** (3), 519-538.

Buell R. W. 2021. Last-Place Aversion in Queues. *Management Science*, **67** (3), 1430-1452.

Bull C., A. Schotter, K. Weigelt. 1987. Tournaments and Piece Rates: An Experimental Study. *Journal of political Economy*, **95** (1), 1-33.

Camerer C. F. 2011. *Behavioral Game Theory: Experiments in Strategic Interaction* Princeton university press.

Charness G., P. Kuhn. 2011. Lab Labor: What Can Labor Economists Learn from the Lab? *Handbook of Labor Economics*. Elsevier, 229-330.

Charness G., U. Gneezy, A. Henderson. 2018. Experimental Methods: Measuring Effort in Economics Experiments. *Journal of Economic Behavior & Organization*, **149**, 74-87.

Chen H., N. Lim. 2013. Should Managers Use Team-Based Contests? *Management Science*, **59** (12), 2823-2836.

Chen H., S. H. Ham, N. Lim. 2011. Designing Multiperson Tournaments with Asymmetric Contestants: An Experimental Study. *Management Science*, **57** (5), 864-883.

Chung K., K. Kim, N. Lim. 2020. Social Structures and Reputation in Expert Review Systems. *Management Science*, **66** (7), 3249-3276.
Davis A. M. 2015. An Experimental Investigation of Pull Contracts in Supply Chains. *Production and Operations Management*, **24** (2), 325-340.
Davis A. M., K. Hyndman. 2018. An Experimental Investigation of Managing Quality through Monetary and Relational Incentives. *Management Science*, **64** (5), 2345-2365.
Davis A. M., V. Gaur, D. Kim. 2021. Consumer Learning from Own Experience and Social Information: An Experimental Study. *Management Science*, **67** (5), 2924-2943.
Davis A. M., E. Katok, N. Santamaría. 2014. Push, Pull, or Both? A Behavioral Study of How the Allocation of Inventory Risk Affects Channel Efficiency. *Management Science*, **60** (11), 2666-2683.
Dechenaux E., D. Kovenock, R. M. Sheremeta. 2015. A Survey of Experimental Research on Contests, All-Pay Auctions and Tournaments. *Experimental Economics*, **18** (4), 609-669.
Delfgaauw J., R. Dur, A. Non, W. Verbeke. 2015. The Effects of Prize Spread and Noise in Elimination Tournaments: A Natural Field Experiment. *Journal of Labor Economics*, **33** (3), 521-569.
Deutscher C., M. Sahm, S. Schneemann, H. Sonnabend. 2019. Strategic Investment Decisions in Multi-Stage Contests with Heterogeneous Players. *Theory and Decision*, 1-37.
Fehr E., G. Kirchsteiger, A. Riedl. 1993. Does Fairness Prevent Market Clearing? An Experimental Investigation. *The quarterly journal of economics*, **108** (2), 437-459.
Fernandez R., D. Rodrik. 1991. Resistance to Reform: Status Quo Bias in the Presence of Individual-Specific Uncertainty. *The American Economic Review*, **81** (5), 1146-1155.
Fischbacher U. 2007. Z-Tree: Zurich Toolbox for Ready-Made Economic Experiments. *Experimental Economics*, **10** (2), 171-178.
Fu Q., J. Lu. 2012. The Optimal Multi-Stage Contest. *Economic Theory*, **51** (2), 351-382.
Harbring C., B. Irlenbusch. 2003. An Experimental Study on Tournament Design. *Labour Economics*, **10** (4), 443-464.
Jiang Z., Y. Huang, D. R. Beil. 2021. The Role of Problem Specification in Crowdsourcing Contests for Design Problems: A Theoretical and Empirical Analysis. *Manufacturing & Service Operations Management*, **23** (3), 637-656.
Kalra A., M. Shi. 2001. Designing Optimal Sales Contests: A Theoretical Perspective. *Marketing Science*, **20** (2), 170-193.
Kim H.-W., A. Kankanhalli. 2009. Investigating User Resistance to Information Systems Implementation: A Status Quo Bias Perspective. *MIS Quarterly*, **33** (3), 567-582.
Kim K., K. Chung, N. Lim. 2019. Third-Party Reviews and Quality Provision. *Management Science*, **65** (6), 2695-2716.
Koh T. K. 2019. Adopting Seekers' Solution Exemplars in Crowdsourcing Ideation Contests: Antecedents and Consequences. *Information Systems Research*, **30** (2), 486-506.
Korpeoglu C. G., E. Körpeoğlu, S. Tunç. 2021. Optimal Duration of Innovation Contests. *Manufacturing & Service Operations Management*, **23** (3), 657-675.
Körpeoğlu E., C. G. Korpeoglu, İ. E. Hafalır. 2022. Parallel Innovation Contests. *Operations Research*, **70** (3), 1506-1530.
Kuziemko I., R. W. Buell, T. Reich, M. I. Norton. 2014. "Last-Place Aversion": Evidence and Redistributive Implications. *The Quarterly Journal of Economics*, **129** (1), 105-149.
Lazear E. P. 1989. Pay Equality and Industrial Politics. *Journal of Political Economy*, **97** (3), 561-580.
Lazear E. P., S. Rosen. 1981. Rank-Order Tournaments as Optimum Labor Contracts. *Journal of Political Economy*, **89** (5), 841-864.
Lim N. 2010. Social Loss Aversion and Optimal Contest Design. *Journal of Marketing Research*, **47** (4), 777-787.
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.
Loewenstein G. F., L. Thompson, M. H. Bazerman. 1989. Social Utility and Decision Making in Interpersonal Contexts. *Journal of Personality and Social Psychology*, **57** (3), 426.

Long X., J. Nasiry, Y. Wu. 2020. A Behavioral Study on Abandonment Decisions in Multistage Projects. *Management Science*, **66** (5), 1999-2016.

Mago S. D., R. M. Sheremeta. 2019. New Hampshire Effect: Behavior in Sequential and Simultaneous Multi-Battle Contests. *Experimental Economics*, **22** (2), 325-349.

Masatlioglu Y., E. A. Ok. 2005. Rational Choice with Status Quo Bias. *Journal of Economic Theory*, **121** (1), 1-29.

Menon N., A. Mishra, S. Ye. 2020. Beyond Related Experience: Upstream Vs. Downstream Experience in Innovation Contest Platforms with Interdependent Problem Domains. *Manufacturing & Service Operations Management*, **22** (5), 1045-1065.

Moldovanu B., A. Sela. 2001. The Optimal Allocation of Prizes in Contests. *The American Economic Review*, **91** (3), 542-558.

Mullainathan S., R. H. Thaler. 2000. Behavioral Economics. Working Paper, NBER.

Nittala L., S. Erat, V. Krishnan. 2022. Designing Internal Innovation Contests. *Production and Operations Management*, **31** (5), 1963-1976.

Northcraft G. B., M. A. Neale. 1987. Experts, Amateurs, and Real Estate: An Anchoring-and-Adjustment Perspective on Property Pricing Decisions. *Organizational behavior and human decision processes*, **39** (1), 84-97.

Orrison A., A. Schotter, K. Weigelt. 2004. Multiperson Tournaments: An Experimental Examination. *Management Science*, **50** (2), 268-279.

Parco J. E., A. Rapoport, W. Amaldoss. 2005. Two-Stage Contests with Budget Constraints: An Experimental Study. *Journal of Mathematical Psychology*, **49** (4), 320-338.

Polites G. L., E. Karahanna. 2012. Shackled to the Status Quo: The Inhibiting Effects of Incumbent System Habit, Switching Costs, and Inertia on New System Acceptance. *MIS Quarterly*, **36** (1), 21-42.

Roels G., X. Su. 2014. Optimal Design of Social Comparison Effects: Setting Reference Groups and Reference Points. *Management Science*, **60** (3), 606-627.

Samuelson W., R. Zeckhauser. 1988. Status Quo Bias in Decision Making. *Journal of Risk and Uncertainty*, **1** (1), 7-59.

Schmitt P., R. Shupp, K. Swope, J. Cadigan. 2004. Multi-Period Rent-Seeking Contests with Carryover: Theory and Experimental Evidence. *Economics of Governance*, **5** (3), 187-211.

Schweitzer M. E., G. P. Cachon. 2000. Decision Bias in the Newsvendor Problem with a Known Demand Distribution: Experimental Evidence. *Management science*, **46** (3), 404-420.

Sheremeta R. M. 2010a. Experimental Comparison of Multi-Stage and One-Stage Contests. *Games and Economic Behavior*, **68** (2), 731-747.

---. 2010b. Expenditures and Information Disclosure in Two-Stage Political Contests. *Journal of Conflict Resolution*, **54** (5), 771-798.

Stracke R., W. Höchtl, R. Kerschbamer, U. Sunde. 2015. Incentives and Selection in Promotion Contests: Is It Possible to Kill Two Birds with One Stone? *Managerial and Decision Economics*, **36** (5), 275-285.

Thaler R. H. 1988. Anomalies: The Winner's Curse. *Journal of economic perspectives*, **2** (1), 191-202.

Thaler R. H. 1999. Mental Accounting Matters. *Journal of Behavioral Decision Making*, **12**, 183-206.

Tullock G. 2001. Efficient Rent Seeking. *Efficient Rent-Seeking*. Springer, 3-16.

Tversky A., D. Kahneman. 1974. Judgment under Uncertainty: Heuristics and Biases: Biases in Judgments Reveal Some Heuristics of Thinking under Uncertainty. *science*, **185** (4157), 1124-1131.

---. 1979. Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, **47** (2), 263-291.

Venkatesh V., A. Rai, T. A. Sykes, R. Aljafari. 2016. Combating Infant Mortality in Rural India: Evidence from a Field Study of Ehealth Kiosk Imlementations. *MIS Quarterly*, **40** (2), 353-380.

Wooten J. O. 2022. Leaps in Innovation and the Bannister Effect in Contests. *Production and Operations Management*, **31** (6), 2646-2663.

Zhang S., P. V. Singh, A. Ghose. 2019. A Structural Analysis of the Role of Superstars in Crowdsourcing Contests. *Information Systems Research*, **30** (1), 15-33.