Information Systems Research

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Salience Bias in Crowdsourcing Contests

Ho Cheung Brian Lee, Sulin Ba, Xinxin Li, Jan Stallaert

INFORMATION SYSTEMS RESEARCH
http://pubsonline.informs.org/journal/isre/
Vol. 29, No. 2, June 2018, pp. 401–418
ISSN 1047-7047 (print), ISSN 1526-5536 (online)

Salience Bias in Crowdsourcing Contests

Ho Cheung Brian Lee,[a] Sulin Ba,[b] Xinxin Li,[b] Jan Stallaert[b]

[a] Manning School of Business, University of Massachusetts Lowell, Lowell, Massachusetts 01854; [b] School of Business, University of Connecticut, Storrs, Connecticut 06269

Contact: brian_lee@uml.edu, http://orcid.org/0000-0002-6971-8454 (HCBL); sulin.ba@uconn.edu (SB); xinxin.li@uconn.edu (XL); jan.stallaert@uconn.edu (JS)

Received: September 1, 2016
Revised: June 21, 2017; December 2, 2017
Accepted: December 30, 2017
Published Online in Articles in Advance: March 7, 2018

https://doi.org/10.1287/isre.2018.0775

Abstract. Crowdsourcing relies on online platforms to connect a community of users to perform specific tasks. However, without appropriate control, the behavior of the online community might not align with the platform's designed objective, which can lead to an inferior platform performance. This paper investigates how the feedback information on a crowdsourcing platform and systematic bias of crowdsourcing workers can affect crowdsourcing outcomes. Specifically, using archival data from the online crowdsourcing platform Kaggle, combined with survey data from actual Kaggle contest participants, we examine the role of a systematic bias, namely, the salience bias, in influencing the performance of the crowdsourcing workers and how the number of crowdsourcing workers moderates the impact of the salience bias on the outcomes of contests. Our results suggest that the salience bias influences the performance of contestants, including the winners of the contests. Furthermore, the number of participating contestants may attenuate or amplify the impact of the salience bias on the outcomes of contests, depending on the effort required to complete the tasks. Our results have critical implications for crowdsourcing firms and platform designers.

History: Panos Constantinides, Ola Henfridsson, and Geoffrey Parker, Special Issue Editors; Mingfeng Lin, Associate Editor. This paper has been accepted for the *Information Systems Research* Special Issue on Digital Infrastructure and Platforms.

Funding: The authors acknowledge the support of the National Natural Science Foundation of China [Project 71229101].

Supplemental Material: The online appendix is available at https://doi.org/10.1287/isre.2018.0775.

Keywords: behavioral economics • crowdsourcing • open innovation • salience bias • parallel path effect • competition effect

1. Introduction

Crowdsourcing contests are widely used as a means for generating ideas and solving problems (e.g., TopCoder, Taskcn, Kaggle) through the hosting of open contests online. Because of the unprecedented scale of the potential workforce for organizations to generate new ideas and innovative solutions (Boudreau and Lakhani 2013), many firms and governmental agencies such as General Electric, NASA, and Procter & Gamble, have used crowdsourcing as part of their research and development (R&D) processes.[1]

As a type of digital platform, a crowdsourcing platform provides flexibility for online users to organically generate new and diverse ideas (Yoo et al. 2012). The diversity and openness are critical to the innovation process (Shin et al. 2012). However, if platform owners do not manage users carefully, such flexibility may cause some users to behave undesirably, which can divert from the platform's objective. This renders the value of the crowdsourcing platforms questionable (Ba et al. 2001, Tilson et al. 2010). Psychologists have shown that human beings tend to be overconfident about the information that can be retrieved easily or is prominent without properly adjusting for the predictive validity or reliability of the information (Tversky and Kahneman 1974, Griffin and Tversky 1992). The researchers give an example where "the impact of seeing a house burning on the subjective probability of such accidents is probably greater than the impact of reading about a [house] fire in the local paper" (Tversky and Kahneman 1974, p. 1127). In other words, humans tend to rely on the information that is explicitly shown to them rather than the information that is presented more implicitly. We refer to this kind of bias as the *salience bias*. This bias has been observed among professionals—who are expected to be rational—in making important business or medical decisions (Camacho et al. 2011, Rudi and Drake 2014). For example, in the case of recruitment screening, a recruiter might focus on the warmth of a reference letter first and then adjust her perception of the letter by considering the credibility of the writer. However, according to Griffin and Tversky (1992), such an adjustment is often insufficient for predicting actual performance. Hence, it creates a salience bias that causes the recruiter to be overconfident about the explicitly presented information (i.e., the warmth of the letter). Research also shows that providing real-time feedback about electricity usage encourages energy conservation because of the salience of the usage information (Tiefenbeck et al. 2018).

On crowdsourcing platforms, contestants receive different sources of information when they participate in crowdsourcing contests, some of which can be more prominent than others. For example, some crowdsourcing platforms provide in-progress feedback to the contestants, which is typically more prominent than other information the contestants have, such as their own evaluation of their solutions. Depending on its design, this type of feedback, however, may not be fully aligned with the goal of the contestants even when it is objectively measured. For example, improving a predictive model in response to the feedback on its performance on only part of the data may lead to overfitting and eventually inferior prediction performance. Yet similar to the above examples, the prominence of this feedback information may induce salience bias among the contestants, causing them to overemphasize the feedback while overlooking other important information that may be helpful for them to achieve desirable outcomes. Such a bias could systematically affect the performance of every worker on the platform, thus becoming detrimental to the functioning of the crowdsourcing platform. It is unclear how crowdsourcing platforms are impacted in such situations and whether they can still perform effectively. This paper adopts the lens of behavioral economics and tournament theory to examine the possible salience bias caused by the feedback information provided during a contest. Our findings have implications on the ability of crowdsourcing platforms to align platform design with the desired platform objective.

Two critical effects contribute to the quality of crowdsourcing outcomes: the parallel path effect (Boudreau et al. 2011) and the competition effect (Boudreau et al. 2016). The parallel path effect refers to a greater likelihood of obtaining a desirable solution when the number of contestants increases; the competition effect suggests that by facing more competitors, high-ability contestants are more likely to be motivated to put more effort into creating new solutions. With the power of the *parallel path effect*, systematic bias in crowdsourcing contests may be reduced by including more contestants with smaller systematic bias. This argument is particularly valid when systematic bias that can negatively affect the performance of a crowdsourcing worker can be attenuated in some way. By contrast, if contestants perceive different feedback information with bias, the *competition effect* would cause the crowdsourcing outcome to deviate further from its objective as a result of high-ability contestants being stimulated to work harder in a particular direction given by the biasedly perceived feedback. This could create an inferior outcome and weaken the power of crowdsourcing contests. From a firm's perspective, investing in an innovation that is subject to systematic bias in crowdsourcing contests may result in inefficient resource management.

To examine how goal-misaligned user behavior can affect the performance of a crowdsourcing platform, we investigate whether crowdsourcing platforms can be immune to individual biases such as the salience bias. Specifically, we address the following three research questions: (1) Does the salience bias influence the performance of the contestants and the outcome of the contests on the crowdsourcing platform? (2) As suggested by the crowdsourcing literature, the competition effect *may amplify* the influence of the salience bias on the outcome of the contests, whereas the parallel path effect *may reduce* the influence of the salience bias. Then, which effect has a greater influence on the impact of the salience bias on the outcome of the contests, and what is the net effect? (3) Are there any other factors that may alleviate the salience bias on crowdsourcing platforms? Studying these questions allows us to identify the potential salience bias that may systematically affect the performance of the contestants, which will lead to inferior overall crowdsourcing platform performance because of the actions of the affected contestants. The results of our study have direct implications for crowdsourcing platform design.

We address these research questions by using an archival data set from Kaggle, a crowdsourcing contest platform for predictive modeling, paired with the results of a survey we conducted among Kaggle contestants. These research questions are important to our understanding of the behaviors of the emerging online community for data science. Very few studies have systematically analyzed the behaviors of data scientists. Yet, anecdotal evidence suggests that the crowdsourcing contestants on Kaggle may overemphasize the feedback information, which is not necessarily fully aligned with the objective of the contestants. For instance, some experienced contestants suggest that relying on the feedback provided by Kaggle could be a bad choice.[2] Other users on Kaggle have also shared similar experiences on discussion forums,[3] suggesting that following the in-progress feedback from Kaggle may lead to a biased solution, and hence a lower final rank. Some users have provided recommendations to the community suggesting that contestants should rely on their own evaluations rather than the feedback. Some of them even suggest that contestants completely ignore the in-progress feedback information.[4] Yet, these phenomena happen not because of the difficulty in understanding the rules of the contests or the inability of the Kaggle contestants to evaluate their solutions. In fact, as shown in Section 6, our survey results suggest that most of the contestants understand very well the rules of the Kaggle contests and how to evaluate their solutions. Therefore, it is likely that the contestants on these crowdsourcing contests are influenced by the salient feedback, which may eventually affect their final performance. As more companies now rely on crowdsourcing contests to access analytics talent, and more crowdsourcing platforms are available for predictive modeling (e.g., Kaggle, CrowdANALYTIX, DrivenData, and Datascience.net), recognizing the potential cognitive biases of the online contestants is important for contest holders and crowdsourcing platforms to design contests in ways that can achieve desirable outcomes.

It is especially intriguing in our research context (i.e., Kaggle contests) whether online contestants can overcome these cognitive biases, since both feedback and outcome are measured based on objective metrics, and data scientists are expected to make the best use of data for optimal decision making. It is therefore interesting to examine whether salience bias plays a significant role in these crowdsourcing contests. Our study also examines different factors that may moderate the effect of salience bias in crowdsourcing contests, such as the number of contestants, reward size, experience, and information cues in contest description. These moderating effects have not been examined in prior studies on salience bias.

Our results provide evidence of the salience bias among crowdsourcing contestants and show that it has a substantial effect on crowdsourcing outcomes. In addition, this salience bias is prevalent, despite the fact that our survey results of Kaggle participants indicate that they recognize the limitations of the feedback information provided during the contests and have the ability to cross-check the usefulness of the information provided. Furthermore, we demonstrate that the impact of the salience bias is amplified when the competition effect dominates, but the impact is mitigated when the parallel path effect dominates. Nevertheless, our results suggest that the salience bias remains persistent among contest winners, regardless of whether the parallel path effect is dominant. These findings have profound implications for both firms (i.e., seekers) and platform designers. In contrast to the suggestions of some economists (e.g., Alevy et al. 2007, List 2011), our results show that competition does not necessarily eliminate individual biases. In other words, systematic bias can still exist at the aggregate level, producing inferior outcomes. Such bias thus reduces the effectiveness of crowdsourcing contests. Furthermore, increasing the number of contestants acts as a double-edged sword—it can enhance both the parallel path effect and the competition effect. Therefore, increasing the number of contestants does not necessarily attenuate systematic bias. Raising workers' awareness of cognitive biases can be more effective in alleviating the impact of salience bias.

We also study two parameters related to the impact of salience bias: contestant experience and contest reward size. Our results suggest that the more experienced the workers are, the less likely it is that they will exhibit the salience bias. This coincides with findings in the literature showing that experienced workers are usually more rational than inexperienced workers (List 2003). Hence, contest holders should consider ways to encourage more experienced contestants to participate in their contests. Finally, we also find that the impact of the salience bias is stronger in contests with larger rewards. This can be explained by the monetary rewards that stimulate contestants to put more effort into developing their solutions in the direction indicated by the biasedly perceived feedback. Hence, we further suggest that overincentivizing can also create an inferior crowdsourcing outcome because of the interaction between reward size and the salience bias. These guidelines may assist platform owners and firms in designing more efficient systems and creating quality crowdsourcing outcomes.

2. Literature Review

2.1. Productivity and Social Welfare of Digital Platforms

The main contribution of this paper is that it shows how the individual behaviors of online users can affect the productivity and effectiveness of a digital platform. Various studies have investigated how the community plays a role in the productivity of such platforms. For instance, the literature examines how the size of a community affects the economics (Rochet and Tirole 2006) and social welfare (Parker and Van Alstyne 2005) of the digital infrastructure and the platforms. When the network size increases, the utility of the agents on two sides of the market also increases (Katz and Shapiro 1985). Although increasing the size of the community may increase demand for and profits of a platform, it may reduce the incentive that drives innovation effort (Boudreau 2012), which might hurt the generativity (Henfridsson and Bygstad 2013) of the platform.

The behavior of a platform community may evolve over time. The incentives for participation and productivity of an online crowd depend on the participation cost, relevancy of the community (Butler et al. 2014), and diversity of the community (Ren et al. 2015). Apart from the size of the community, individual biases may also be detrimental to the social welfare of digital platforms. For example, studies have shown that mistrust among users can reduce resource allocation efficiency on crowdfunding platforms (Burtch et al. 2014, Lin and Viswanathan 2016). However, the problems caused by this type of mistrust could be solved by an online feedback system (Ba 2001, Dellarocas 2003). This paper extends the literature by elucidating how individual biases can lower or raise the productivity of digital platforms.

2.2. User Behavior in Crowdsourcing Contests

This paper also contributes to the crowdsourcing literature by analyzing how the behavioral anomalies of crowdsourcing workers jeopardize the crowdsourcing outcome. Currently, theories in the crowdsourcing literature focus on how the solution quality is affected by contest parameters such as the number of contestants (Terwiesch and Xu 2008; Boudreau et al. 2011, 2016), reward size (Liu et al. 2014), experience and expertise of the contestants (Jeppesen and Lakhani 2010, Bayus 2013, Huang et al. 2014), the perceived ease of use of the platform (Blohm et al. 2016), and the structure of the crowdsourcing teams (Dissanayake et al. 2015). However, none of these studies address the behavioral anomalies of crowdsourcing workers or how these anomalies can affect the solution outcomes of crowdsourcing contests. According to our research, only Liu et al. (2014) investigated behavioral anomalies on crowdsourcing platforms. They showed that high-ability workers tend to avoid entering a crowdsourcing contest that already has a high-ability incumbent. Our paper investigates a different type of cognitive bias that can be applied to everyone on the platform, not only to the high-ability contestants. Hence, the overall performance of the crowd could be affected systematically. To our knowledge, there is no existing research focusing on how systematic biases affect the outcomes of a crowdsourcing platform.

Furthermore, a number of studies in the area of crowdsourcing contests are grounded in tournament theory. In particular, Connelly et al. (2014) identified a research gap regarding how cognitive bias can affect tournament outcomes. Among all of the findings in the literature on tournament theory, the most studied systematic biases in tournament applications are risk-related (Becker and Huselid 1992, Pope and Schweitzer 2011, Berger and Pope 2011). Our study contributes to this research stream by clarifying how a specific cognitive bias—the salience bias, which changes how workers perceive feedback information—can also influence tournament outcomes.

2.3. In-Progress Feedback in Crowdsourcing Contests

There have been a few studies that have examined the impact of in-progress feedback on crowdsourcing outcomes. Yang et al. (2009) discussed the role of feedback in crowdsourcing contests but did not empirically examine the impact of feedback on the contestants. Adamczyk et al. (2011) combined a survey and a clustering analysis to study how contestants use feedback differently in a crowdsourcing contest. However, their paper did not investigate the mechanism by which feedback affects contestants' performance. Bockstedt et al. (2016) argued that crowdsourcing contestants who submit earlier have a better chance of winning because their solutions are more likely to be evaluated carefully by the seekers, who can evaluate solutions more effectively when only a few solutions are submitted.

Most relevant to our study, Wooten and Ulrich (2017) discussed the role of feedback in a crowdsourcing contest for logo design using a field experiment. They suggested that providing in-progress feedback to the contestants may lead to less variation in submissions because the contestants are likely to tailor their submissions toward the seeker's preferences revealed in the feedback. In their context, following the feedback is rational because the feedback reveals the seekers' preferences, which are directly linked to how the winners will be selected. By contrast, our paper studies a cognitive bias that likely leads to an inferior outcome when the contestants overemphasize the in-progress feedback that may not be aligned with the objective of the contestants, that is, winning the contest.

Broadly speaking, the difference between our paper and Wooten and Ulrich's (2017) is also a key difference between our study and almost all prior studies that examine the role of in-progress feedback in crowdsourcing contests; that is, in prior studies, the feedback is typically based on subjective opinions of seekers who are also the ones selecting the final winners. Therefore, it is natural and rational for the contestants to follow the feedback from the seekers because it reveals the seekers' preferences. By contrast, in our research context, the final performance is evaluated based on objective measures and following the feedback can lead to inferior outcomes. Yet, the contestants still overemphasize the feedback because of their cognitive bias. Such effect of cognitive bias has not been examined in this stream of research. We contribute to this stream of research by showing that in-progress feedback—even if it is objectively measured—may still be misaligned with the objective of the contestants (i.e., winning the contest) and induce salience bias among the contestants. This deepens our understanding of how in-progress feedback may affect the outcomes of crowdsourcing contests.

3. Study Context: Kaggle

Kaggle is a platform that allows firms (or seekers) to post their data and describe problems to seek solutions. Most of these problems are hosted as contests, for which seekers can define their own rules. The contestants on the platform can view the rules before deciding whether to participate or not. Each contest winner typically receives a reward, which might be money, a work opportunity, or an opportunity to attend a conference.

The most common contests on Kaggle are predictive modeling problems. In these contests, the seeker provides a training set and a test set to the contestants. The training set contains both inputs and outputs for the contestants to train their models, whereas the test set contains only the input data. The objective for the contestants is to use their trained models to predict the output of the test set (which is not visible to them). The contestants are allowed multiple uploads of

Figure 1(a). (Color online) Public Leaderboard (Visible to Every Contestant)

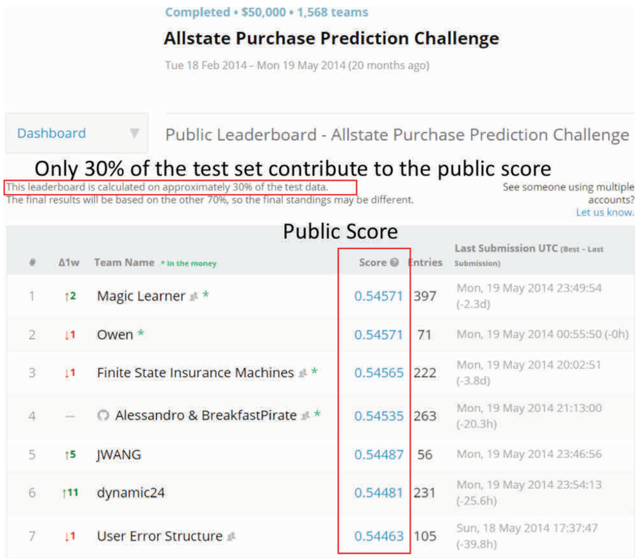
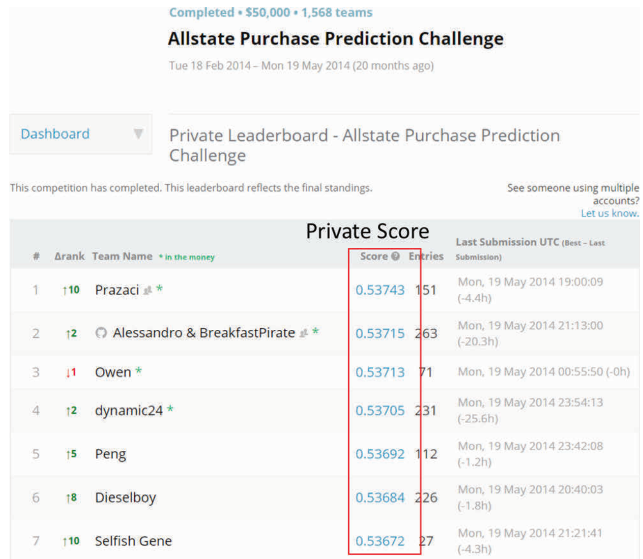


Figure 1(b). (Color online) Private Leaderboard (Hidden from the Contestants Until the Contest Is Over)



the solutions they derive from the test set during the learning phase. Every time a contestant uploads her solution, the system shows the score of the uploaded solution to the contestant; this score is called the *public score*. The highest public score of the contestant will be posted on a public leaderboard (Figure 1(a)) that is viewable by every contestant.

The public scores do not determine the final ranking of the contestants—they are merely performance indicators of the effectiveness of the contestants' solutions and are based on only a certain proportion of the test set. (The contestants are not informed which proportion of the test set the scores are derived from.) As shown at the top of the leaderboard in Figure 1(a), the public scores were calculated from 30% of the test set.

By the end of the contest, the contestants select the solutions that can maximize their chances of winning based on their own evaluations and the public scores of the solutions. This process is known as the *submission phase*. After the contest is over, the final ranking of the contestants (private leaderboard) is announced, which is determined by their *private scores*. The private score evaluates a solution in the same way as the public score, except that it uses only the proportion of the test set that is not used in calculating the public score. For example, the private scores shown in Figure 1(b) were computed using the remaining 70% of the test set. Neither the private scores nor the private leaderboard is revealed to the contestants until the contest is over.

4. Hypothesis Development

During the contest, the contestants develop solutions using the training set in the learning phase and select solutions in the submission phase to maximize their chances of winning. In the learning phase, the contestants should evaluate their models by predicting their out-of-sample performance on the test set. A common practice is to cross-validate a model to obtain a generalized predictive model (i.e., a predictive model that minimizes the out-of-sample errors). For example, cross-validation (e.g., *k*-fold cross-validation) is part of the curriculum of many popular machine learning online courses.[5] Such a practice often leads to a relatively accurate estimation of the out-of-sample errors. Theoretically, the expectation of the sample errors obtained through cross-validation is the same as the expectation of the true out-of-sample errors (Abu-Mostafa et al. 2012, p. 139). Hence, contestants in predictive modeling contests should be able to evaluate their own solutions without any feedback from the platform by using statistical methods such as cross-validation on the training set. We refer to this as the *internal evaluation* of the solutions by the contestants themselves.

When the contestants upload their solutions, their public scores provide another, more salient type of feedback on how well their solutions perform, in addition to their internal evaluations. As discussed in Section 1, many studies have shown that the salience bias affects the decisions of individuals, including professionals (e.g., Camacho et al. 2011, Rudi and Drake 2014), and anecdotal evidence shows that the salience bias is also common in the Kaggle community.

In predictive modeling contests, contestants are ideally supposed to improve the accuracy of their out-of-sample predictions when developing their solutions. If the solutions are not biased to any data set, the out-of-sample prediction errors should be similar for any out-of-sample data set. Formally, we denote the score function $s(Z, A)$ for the prediction accuracy of algorithm A given data set Z. The more accurately A predicts Z, the higher the score $s(Z, A)$ will be. Given two test sets Z^1 and Z^2, which are drawn from the same population, we expect that $E(s(Z^1, A)) = E(s(Z^2, A))$. Similarly, if the data sample for the public score is Z^{Public} and the data sample for the private score is Z^{Private}, when both Z^{Public} and Z^{Private} are drawn from the same population, the difference between the public and private scores when A is unbiased is expected to be zero: $E(\Delta S) = 0$, where $\Delta S = s(Z^{\text{Public}}, A) - s(Z^{\text{Private}}, A)$. However, if the salience effect of the public scores causes any contestant to intentionally train their model mainly according to the feedback Z^{Public} while deemphasizing an internal evaluation based on cross-validation (a practice known as overfitting), then the predictive error for Z^{Public} will be smaller than the predictive error for Z^{Private}. Hence, $E(\Delta S) > 0$; that is, over time, the public scores will be systematically higher than the private scores in the presence of the salience bias.

When the salience bias is present, the difference between public and private scores will also likely increase as a contestant receives more feedback. Every time a contestant uploads a solution online, the contestant receives the public score for this solution. If there is no salience bias toward the public score feedback, the contestant should not favor any portion of the test set when developing future solutions, so the public and private scores should increase by a similar magnitude if there is any improvement in future solutions, and the score difference should not be statistically different from its expected mean (which is zero). However, if there is salience bias, the more public scores the contestant receives, the more salient the public score feedback will be, and the more likely it is that the contestant will improve the next solution in the direction of getting a better public score, that is, the more likely it is that the contestant will *overfit* the portion of the test set for deriving public scores when developing future solutions. Therefore, we hypothesize the following: the more feedback a contestant receives, the more likely it is that the public score of the next solution developed by the contestant will be higher than its private score. In other words, the score difference (i.e., $E[\Delta S]$) increases as the contestants receive more feedback because of the salience bias. Accordingly, we formulate our first hypothesis as follows:

Hypothesis 1. *Because of the salience bias, public scores are systematically greater than private scores, and this difference increases as contestants receive more feedback during the learning phase.*

The salience bias can appear in both the learning and submission phases. In the submission phase, the contestants reevaluate their models by predicting their out-of-sample performance on the test set to select which models to submit as their final solutions. Similar to the learning phase, the contestants should be able to evaluate their own solutions without any feedback from the platform by using statistical methods such as cross-validation with the training set. We can thus follow the same logic as that used in the learning phase to argue that, ideally, if the contestants are not biased toward any portion of the test set, the out-of-sample prediction errors should be the same for any portion of the out-of-sample test set; therefore, the chance of a solution with a higher public score being selected in the submission phase should match that of a solution with a higher private score being selected. However, in the presence of a salience bias, the contestants may rely on the public scores more than their internal evaluations, which biases them toward solutions with higher public scores (i.e., the solutions that overfit the portion of the test set for deriving the public scores). Therefore, we can observe a similar pattern in the submission phase: because of the salience bias, solutions with high public scores are more likely to be selected as final solutions than solutions with high private scores.

Hypothesis 2. *Because of the salience bias, solutions with high public scores are more likely to be selected as final solutions than solutions with high private scores during the submission phase.*

Whereas Hypotheses 1 and 2 relate to the presence of a salience bias at the individual contestant level, this salience bias may or may not carry over to the aggregate level of the crowd and affect the final outcomes as well; that is, another crucial question we are investigating is whether the outcomes of crowdsourcing contests can be immune to the salience bias. Because the quality of a contest outcome is ultimately determined by the quality of the solutions submitted by the winners, if the winners are unaffected by the salience bias, the seeker can still receive the desired outcome from the contest. However, because a salience bias is likely to systematically affect every contestant—including those who eventually win the contest—it may not necessarily be possible to completely eliminate the negative impact of the salience bias. Hence, we test whether Hypotheses 1 and 2 also apply to the prize winners, that is, whether the salience bias affects the winners as well.

Hypothesis 3. *A salience bias exists among the prize winners of a crowdsourcing contest in both the learning and submission phases.*

We now examine how the parallel path effect and competition effect moderate the influence of the salience bias on the winners. When there is a sufficient number of contestants, both the parallel path effect (Boudreau et al. 2011) and competition effect (Boudreau et al. 2016) are at work. These two effects differ in how they influence the impact of the salience bias. The parallel path effect mainly pertains to the quality of the winning solutions, whereas the competition effect is mostly about the effort of the high-ability contestants (e.g., winners in our context).

According to the first order statistics, the chance of getting a high-ability contestant will increase as more contestants participate; that is, if the ability of a contestant is a random variable X, and given that there are n contestants, the chance of getting at least one contestant with an ability higher than a threshold x will be $p = 1 - [\Pr(X \leq x)]^n$. Hence, it is more likely to get a high-ability contestant, who is less influenced by the salience bias, to be the winner when there are more contestants (i.e., p increases as n increases). In the crowdsourcing literature, this is also known as the parallel path effect (Boudreau et al. 2011). Although increasing the number of competitors in a contest may also reduce the overall incentives of the competitors to exert effort, Boudreau et al. (2011) suggest that the parallel path effect dominates the overall effort-reducing effect for more uncertain problems. Since predictive modeling problems are typically high-uncertainty problems and can benefit from having a diverse workforce to look for the best approach or path to a solution, the parallel path effect should dominate the overall effort-reducing effect in our context.

When high-ability contestants face more competitors, they tend to work harder to maintain their chances of winning. This phenomenon is known as the competition effect (Boudreau et al. 2016). The competition effect applied to our context suggests that a higher number of contestants leads to high-ability contestants exerting stronger effort in improving and evaluating their solutions (e.g., according to the public scores and/or their internal evaluations). When a salience bias is at work, however, such contestants can expend more effort under the guidance of salient information and are thus more likely to place a greater emphasis on the public scores (Hypothesis 1). Consequently, winners may be subjected to a stronger salience bias than they would be in a contest with few contestants.

During the learning phase, both the parallel path effect and competition effect are at work. Therefore, we develop two competing hypotheses about the salience bias in the learning phase:

Hypothesis 4A (Dominating Parallel Path Effect). *During the learning phase, as the number of contestants increases, the impact of the salience bias on the winners' solutions is reduced.*

Hypothesis 4B (Dominating Competition Effect). *During the learning phase, as the number of contestants increases, the impact of the salience bias on the winners' solutions is further amplified.*

There are two main tasks in predictive modeling contests: creating new solutions and evaluating existing solutions. During the learning phase, contestants have already exerted effort improving and evaluating their solutions. Hence, during the submission phase, the contestants must exert considerably less effort[6] when selecting the solutions compared with the effort required during the learning phase. The parallel path effect should thus dominate the competition effect during the submission phase. Therefore, we posit the following hypothesis:

Hypothesis 5. *During the submission phase, as the number of contestants increases, the influence of the salience bias on the winners' solutions is attenuated.*

5. Model

This section introduces two models to address the proposed hypotheses. We test our hypotheses by examining (a) the hypothesized patterns in the score differences and (b) the types of solutions that are more likely to be selected as the final solutions. Table 1 lists the definitions of the main variables.

5.1. Learning Phase Model

As suggested in Section 4, the salience bias can systematically lead to higher public scores than private scores if the contestants train their models mainly according to the feedback received (i.e., the public score) while deemphasizing cross-validation.

Formally, the public and private scores of the jth solution uploaded by contestant i in each contest c are denoted as S_{ijc}^{Public} and S_{ijc}^{Private}, respectively. The term S_{ijc}^{Public} (S_{ijc}^{Private}) measures the prediction accuracy of an algorithm created by contestant i for solution j in contest c when applied to the data sample contributing to the public score (private score) calculation in contest c. The score difference is defined as the difference between the public and private scores; that is, $\Delta S_{ijc} = S_{ijc}^{\text{Public}} - S_{ijc}^{\text{Private}}$.

An intuitive approach for testing whether the salience bias exists during the learning phase is to determine whether the score difference is systematically greater than zero (i.e., $E[\Delta S_{ijc}] > 0$) and how ΔS_{ijc} changes as the number of feedback scores increases. However, Kaggle allows each seeker to use different measurements in evaluating the prediction accuracy of their solutions. To be consistent across all contests, ΔS_{ijc} is standardized for each contest by (a) switching the signs of the scores that are aimed to be minimized, so that higher scores always indicate an improvement, and (b) normalizing the magnitude of ΔS_{ijc} for each contest, so that the magnitudes of ΔS_{ijc} are comparable among all contests. In particular, some evaluation

Table 1. Definition of the Main Variables

Main variable	Description
$\widetilde{\Delta S}_{ijc}$	The standardized score difference of the jth solution uploaded by contestant i in contest c
$Selected_{ijc}$	The indicator variable that shows whether solution j of contestant i is chosen by the contestant to submit in contest c as a final solution
$\#Feedback_{ijc}$	The number of feedback scores that contestant i receives before uploading solution j in contest c
$Time_{ijc}$	The time elapsed (in days) when contestant i uploaded solution j in contest c starting from the time when the contestant uploaded her first solution in contest c divided by her total time available in contest c
$\widetilde{RankInTeam}^{\text{Public}}_{ijc}$	The normalized final rank based on the public score of solution j among other solutions of contestant i in contest c
$\widetilde{RankInTeam}^{\text{Private}}_{ijc}$	The normalized final rank based on the private score of solution j among other solutions of contestant i in contest c
$Winner_{ic}$	The indicator variable that shows whether contestant i is a winner of contest c
$\#Team_c$	The total number of contestants in contest c

scores fall in the range between 0 and 1 (e.g., area under curve), whereas others have no upper bound (e.g., root mean squared error). We construct the standardized score difference of each solution j of contestant i in contest c as follows:

$$\widetilde{\Delta S}_{ijc} = \frac{\Delta S_{ijc}}{\sigma_c(\Delta S_{ijc})}, \tag{1}$$

where $\sigma_c(\Delta S_{ijc})$ is the sample standard deviation of ΔS_{ijc} in contest c. Theoretically, if the contestants do not overfit their models to public scores, as suggested in Section 4, the expected difference between the public and private scores (i.e., the mean of ΔS_{ijc}) should be zero, and hence $E[\widetilde{\Delta S}_{ijc}]$ should also be zero.

To examine Hypothesis 1, we first investigate the sign of $\widetilde{\Delta S}_{ijc}$ using the Wilcoxon signed rank test as a nonparametric test for pairwise comparisons. To study whether $\widetilde{\Delta S}_{ijc}$ increases with the number of feedback scores, we construct a linear regression model in which $\widetilde{\Delta S}_{ijc}$ is the dependent variable and use the number of feedback scores ($\#Feedback_{ijc}$) received by contestant i before the contestant uploads the jth solution in contest c as an independent variable. We perform a logarithmic transformation on the variable $\#Feedback_{ijc}$ to correct its skewed distribution (Gelman and Hill 2007). Thus, the following baseline *learning phase model* is derived:

$$\widetilde{\Delta S}_{ijc} = \gamma_i + \delta_c + \alpha_1 Time_{ijc} + \alpha_2 \log(\#Feedback_{ijc}) + \epsilon_{ijc}, \tag{2}$$

where α_1 and α_2 are the coefficients, and ϵ_{ijc} is the error term. To control for any potential trend in score differences over time, we include the variable $Time_{ijc}$, which is computed as the time elapsed when contestant i uploaded solution j in contest c starting from the time when the contestant uploaded her *first* solution in contest c, divided by her *total* time available in contest c.[7] We also include the fixed effects for contest c (δ_c) and contestant i (γ_i) to control for the idiosyncratic characteristics of each contest and contestant. These fixed effects also contribute to absorbing the systematic difference in the scoring criteria among different contests. If Hypothesis 1 is valid, the estimated α_2 should be positive, that is, $\alpha_2 > 0$. This implies that the score difference increases as the contestants receive more feedback.

5.2. Submission Phase Model

In the submission phase of each contest, Kaggle requires the contestants to select final solutions because the contestants tend to upload more solutions than the contest requires during the learning phase. To test Hypothesis 2, we examine whether solutions with high public scores are more likely to be selected than those with high private scores. To avoid measurement inconsistency among the contests, instead of including the actual public and private scores in the model, we use the rankings of the public and private scores relative to those of other solutions uploaded by the same contestant in the same contest (i.e., $RankInTeam^{\text{Public}}_{ijc}$ and $RankInTeam^{\text{Private}}_{ijc}$). Because the ranking of the solutions also depends on how many solutions a contestant has uploaded, we normalize the score ranks as follows:

$$\widetilde{RankInTeam}^{\text{Public}}_{ijc} = 1 - \frac{RankInTeam^{\text{Public}}_{ijc}}{J_{ic}}; \tag{3}$$

$$\widetilde{RankInTeam}^{\text{Private}}_{ijc} = 1 - \frac{RankInTeam^{\text{Private}}_{ijc}}{J_{ic}}. \tag{4}$$

Here, J_{ic} is the total number of solutions uploaded by contestant i in contest c. We model the likelihood of different solutions being selected by using a conditional logit model. Assume each solution j uploaded by contestant i in contest c has a latent value V_{ijc}. The likelihood of solution j being selected depends on the magnitude of V_{ijc}, which can be represented as a function of both public scores and private scores, as follows:

$$V_{ijc} = \gamma_i + \delta_c + \beta_1 Time_{ijc} + \beta_2 \widetilde{RankInTeam}^{\text{Public}}_{ijc} + \beta_3 \widetilde{RankInTeam}^{\text{Private}}_{ijc} + \varepsilon_{ijc}, \tag{5}$$

where β_1, β_2, and β_3 are the coefficients, and ε_{ijc} is the error term that follows a Gumbel (type 1 extreme value) distribution (Greene 2008). The fixed effects (γ_i and δ_c) are included to control for the idiosyncratic characteristics of each contestant and contest.

We use the dummy variable $Selected_{ijc}$ to indicate whether solution j is selected by contestant i in contest c as a final solution. Let $\mathbf{Selected}_{ic} = (Selected_{i1c}, \ldots, Selected_{iJ_{ic}c})$ capture all of the selected solutions by contestant i in contest c. A closed form model can then be formulated as derived by Hosmer et al. (2013)

$$\Pr\left(\mathbf{Selected}_{ic} \,\middle|\, \sum_{j=1}^{J_{ic}} Selected_{ijc} = K_c\right) = \frac{\exp(\sum_{j=1}^{J_{ic}} Selected_{ijc} \cdot X_{ijc}\beta)}{\sum_{\mathbf{d}_{ic}\in D_{ic}} \exp(\sum_{j=1}^{J_{ic}} d_{ijc} \cdot X_{ijc}\beta)}, \tag{6}$$

where X_{ijc} is a row vector consisting of the solution-variant variables (i.e., the variables in Model (5) excluding the fixed effects[8]), β is the corresponding coefficients column vector, K_c is the number of solutions that can be selected in contest c, $\mathbf{d}_{ic}$ is a vector containing dummy variables d_{ijc} with $\sum_{j=1}^{J_{ic}} d_{ijc} = K_c$, and D_{ic} is a set containing all possible $\mathbf{d}_{ic}$'s that satisfy the condition $\sum_{j=1}^{J_{ic}} d_{ijc} = K_c$. Hypothesis 2 is supported if $\beta_2 > \beta_3$, that is, the solutions with high public scores are more likely to be selected than the solutions with high private scores.

5.3. Winner Performance

Because the winners define the outcomes of crowdsourcing contests, we examine whether the winners exhibit a stronger or weaker salience bias than the other contestants. To examine Hypothesis 3, we first investigate the sign of $\widetilde{\Delta S}_{ijc}$ for the solutions uploaded by contest winners using the Wilcoxon signed rank test. We then identify the winners of the contests (the contestants who received rewards) with the indicator variable $Winner_{ic}$, where $Winner_{ic} = 1$ when contestant i is a winner of the contest c, and include it in our regression models. The main effect of $Winner_{ic}$ is absorbed by the contestant fixed effect. Thus, Models (2) and (5) are modified as Models (2′) and (5′), respectively, as follows:

$$\widetilde{\Delta S}_{ijc} = \gamma_i + \delta_c + \alpha_1 Time_{ijc} + \alpha_2 \log(\#Feedback_{ijc}) + \alpha_3 Winner_{ic} \times \log(\#Feedback_{ijc}) + \epsilon_{ijc}; \tag{2′}$$

$$V_{ijc} = \gamma_i + \delta_c + \beta_1 Time_{ijc} + \beta_2 \widetilde{RankInTeam}^{\text{Public}}_{ijc} + \beta_3 \widetilde{RankInTeam}^{\text{Private}}_{ijc} + \beta_4 Winner_{ic} \times \widetilde{RankInTeam}^{\text{Public}}_{ijc} + \beta_5 Winner_{ic} \times \widetilde{RankInTeam}^{\text{Private}}_{ijc} + \varepsilon_{ijc}. \tag{5′}$$

If Hypothesis 3 is true, then the coefficient of *#Feedback* should remain positive for the winners in the learning phase model (Model (2′)); that is, $\alpha_2 + \alpha_3 > 0$. Similarly, the solutions with high public scores should still have a higher likelihood of being selected than the solutions with high private scores in the submission phase model (Model (5′)); that is, $\beta_2 + \beta_4 > \beta_3 + \beta_5$. In Section 6.5, we also provide a subsample analysis, which estimates the model by using only the winners.

5.4. Moderating Effects of the Parallel Path Effect and Competition Effect

To test Hypotheses 4A, 4B, and 5, we include the variables related to the total number of contestants. In both phases, we use the total number of contestants in contest c ($\log(\#Team_c)$) to capture how competitive the contest is as well as the magnitude of the parallel path effect. A log transformation is used to correct the skewness of the variable (Gelman and Hill 2007). Its main effect is absorbed by the contest fixed effect. Hence, the learning phase and submission phase models are modified as Models (2″) and (5″), respectively

$$\begin{aligned}\widetilde{\Delta S}_{ijc} = \gamma_i + \delta_c &+ \alpha_1 Time_{ijc} + \alpha_2 \log(\#Feedback_{ijc}) \\ &+ \alpha_3 Winner_{ic} \times \log(\#Feedback_{ijc}) \\ &+ \alpha_4 \log(\#Team_c) \times \log(\#Feedback_{ijc}) \\ &+ \alpha_5 Winner_{ic} \times \log(\#Team_c) \times \log(\#Feedback_{ijc}) \\ &+ \epsilon_{ijc};\end{aligned} \tag{2″}$$

$$\begin{aligned}V_{ijc} = \gamma_i + \delta_c &+ \beta_1 Time_{ijc} + \beta_2 \widetilde{RankInTeam}^{\text{Public}}_{ijc} \\ &+ \beta_3 \widetilde{RankInTeam}^{\text{Private}}_{ijc} \\ &+ \beta_4 Winner_{ic} \times \widetilde{RankInTeam}^{\text{Public}}_{ijc} \\ &+ \beta_5 Winner_{ic} \times \widetilde{RankInTeam}^{\text{Private}}_{ijc} \\ &+ \beta_6 \log(\#Team_c) \times \widetilde{RankInTeam}^{\text{Public}}_{ijc} \\ &+ \beta_7 \log(\#Team_c) \times \widetilde{RankInTeam}^{\text{Private}}_{ijc} \\ &+ \beta_8 Winner_{ic} \times \log(\#Team_c) \times \widetilde{RankInTeam}^{\text{Public}}_{ijc} \\ &+ \beta_9 Winner_{ic} \log(\#Team_c) \times \widetilde{RankInTeam}^{\text{Private}}_{ijc} \\ &+ \varepsilon_{ijc}.\end{aligned} \tag{5″}$$

If the parallel path effect dominates the competition effect (Hypothesis 4A) during the learning phase, the marginal effect of the public score feedback among the winners should be reduced when the number of contestants increases; that is, $\alpha_4 + \alpha_5$ should be negative in Model (2″). By contrast, if the competition effect dominates (Hypothesis 4B), the marginal effect of the feedback should be amplified when the number of contestants increases; hence, $\alpha_4 + \alpha_5$ should be positive in Model (2″). Similarly, if a parallel path effect is at work, then, according to Hypothesis 5, during the submission phase, the solutions with high public scores will be less likely to be selected by the winners as the number of contestants increases, and/or the solutions with high private scores will be more likely to be selected by the winners as the number of contestants increases;

that is, $\beta_6+\beta_8$ should be negative and/or $\beta_7+\beta_9$ should be positive in Model (5'').

6. Data and Empirical Analysis

Section 5 presents two models: the learning phase model and the submission phase model. In this section, we illustrate how the model coefficients are estimated. First, we describe the archival data from Kaggle used to estimate the model. Then, we provide some model-free evidence for the hypothesized patterns in the presence of salience bias. Next, we present our empirical results from the Kaggle data and discuss the identification issue based on the results of a survey conducted among Kaggle contestants. Finally, we perform robustness tests.

6.1. Data

In this study, we use an archival data set from Kaggle. The data set contains 239 completed contests before September 1, 2015. Because this study investigates how a salience bias arises from the impact of feedback information (i.e., the public scores), we limit our scope by examining only those contests that implemented a public leaderboard. We also exclude those where overfitting was inapplicable (i.e., the contests were unrelated to predictive modeling). Furthermore, contestants may differ in how they value nonmonetary rewards. For example, some contests were hosted as classroom practice for machine learning classes, and other contests provided points on Kaggle instead of a monetary reward. To be consistent, we use only those contests that offered monetary rewards. In total, among 239 contests in our data set, we identify 103 predictive modeling contests that provided monetary rewards and implemented a public leaderboard. (Most of the other contests were in-class competitions.) Table 2 presents a descriptive summary of the main variables.

Our sample comprises 44,827 teams with a total of 258 winning spots among all contests. Overall, the sample comprises 695,622 observations at the solution level. In each contest, we observe rule parameters including the reward size, number of winners, and percentage of the test set that contributed to the public score. We also observe every contestant submission, its public score and private score,[9] the time of upload, the contestant's name, and whether the submission was selected by the contestant as a final solution.

Table 2. Descriptive Summary of the Main Variables

Main variable	Mean	St. dev.	No. of observations
$\widetilde{\Delta S}_{ijc}$	0.226	1.314	695,622
$Selected_{ijc}$	0.144	0.351	695,622
$\widetilde{\#Feedback}_{ijc}$	32.993	49.423	695,622
$\widetilde{RankInTeam}^{\text{Public}}_{ijc}$	0.472	0.292	695,622
$\widetilde{RankInTeam}^{\text{Private}}_{ijc}$	0.475	0.294	695,622
$Time_{ijc}$	0.557	0.333	695,622
$Winner_{ic}$	0.006	0.076	44,827
$\#Team_c$	435.214	551.023	103

In each contest, a contestant can participate on her own, or several contestants can participate together as a team. A team, once formed, stays the same throughout the contest. In this paper, we use "team" and "contestant" interchangeably. If several people participate as a team, they are treated no differently than a single contestant, so they are expected to collaborate as a team to work toward the same goal (i.e., to win the contest). In our data set, the average number of members on each team is approximately 1.15, suggesting that the majority of teams have only one contestant. To prevent collusion among the contestants, Kaggle removes users who submit from multiple accounts or privately share between teams. The number of winning spots in a contest is typically between one and three and is predetermined before the contest starts.

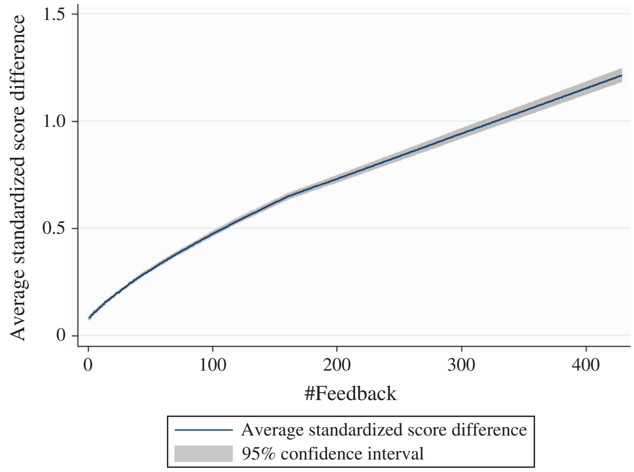
6.2. Model-Free Evidence

We start by exploring whether our hypothesized patterns for public and private scores in the presence of salience bias show in our data. First, we conduct the Wilcoxon signed rank test to compare the public scores and private scores. The test results reveal that public scores are statistically larger than private scores at a 1% significance level for all solutions, including the solutions uploaded by the winners. Thus, the sign of the score difference hypothesized in Hypotheses 1 and 3 is supported.

Second, we explore the relationship between the score difference and the number of feedback scores by first dividing the solutions into 50 equal-sized groups based on the number of feedback scores.[10] We then calculate the average standardized score difference and 95% confidence interval for each group of solutions, and plot their relationship with the number of feedback scores, shown in Figure 2. It shows that the score difference increases as more feedback is received, again consistent with the pattern hypothesized in Hypothesis 1.

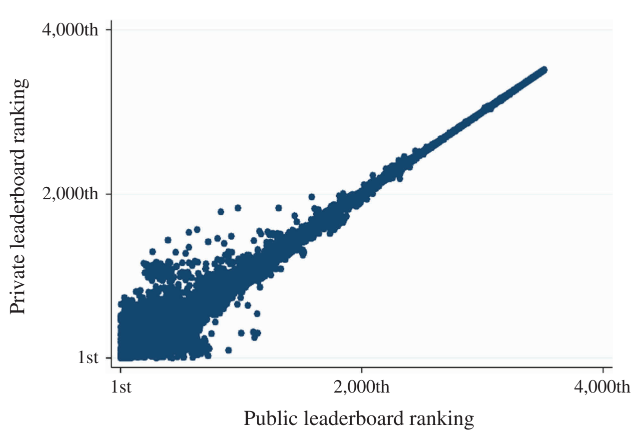
Third, we examine whether public score ranking and private score ranking are aligned, that is, whether the solutions ranked better on the public leaderboard are also ranked better on the private leaderboard. As shown in Figure 3, the ranking correspondence becomes insignificant as we move to the top of the leaderboard (the lower left corner of the graph); that is, the public leaderboard ranking becomes less representative of the final ranking as the public score ranking improves. This could be the result of contestants who are ranked best on the public leaderboard being the ones overfitting the public scores and therefore performing worse in the final ranking (i.e., on the private leaderboard).

Figure 2. (Color online) Score Difference vs. Number of Feedback Scores



In Table 3, we also check whether the public score ranking and the private score ranking are aligned by examining their correlations. As we move up the public leaderboard, the correlation between the public score ranking and the private score ranking goes down and eventually becomes close to zero and insignificant. If we move up the private leaderboard, it shows a similar pattern, but the correlation numbers are generally higher than if we move up the public leaderboard. This suggests that compared to the contestants with the same position on the public leaderboard, the contestants who are ranked best on the private leaderboard are more likely to be at a similar rank on the public leaderboard; that is, the contestants who paid more attention to the final performance are more likely to do well in both rankings than the contestants who are ranked high on the public leaderboard, likely because of salience bias and the overfitting tendency by the latter.

Figure 3. (Color online) Illustration of Ranking Correspondence



6.3. Regression Results

The estimation results of our regression models are shown in Tables 4 and 5. Robust standard errors clustered at the contestant level are reported in parentheses.[11] In both tables, column (1) presents the baseline models and shows evidence of the salience bias. The differences between the winners and other contestants are shown in column (2), and the moderating effects of the parallel path effect and competition effect are shown in column (3).

Our results suggest that a salience bias is persistent among contestants in both phases. In column (1) of Table 4, the positive coefficient of log(#*Feedback*$_{ijc}$) ($\alpha_2 = 0.05633$) indicates that a 1% increase in the number of feedback scores would predict an increase of 0.0005633 units of the standardized score difference, which is about 0.25% of the mean of the standardized score difference (the mean is 0.226 in Table 2). Thus, Hypothesis 1 is supported, suggesting that the effect of salience bias is stronger for the contestants who receive more feedback.

Similarly, column (1) in Table 5 supports Hypothesis 2 because the difference between the coefficients of $RankInTeam^{Public}_{ijc}$ and $RankInTeam^{Private}_{ijc}$ is positive at the 1% significance level ($\beta_2 > \beta_3$). This suggests that a solution with a high public score is more likely to be selected than a solution with a high private score.

Notably, in column (1) of Table 5, the private scores have a tangible influence on which solutions the contestants select. Although the private scores are not directly observed by the contestants during the contest, the statistical significance of the private scores likely captures the effect of cross-validation. As shown in Abu-Mostafa et al. (2012), the expected value of the error terms for cross-validation approximates the expected value of the out-of-sample error. Since the private score basically represents the value of the out-of-sample error, private scores can be considered proxies for the cross-validation scores that the contestants obtain for their solutions.

In column (2) of Tables 4 and 5, we test Hypothesis 3 for the learning phase (i.e., H_0: $\alpha_2 + \alpha_3 \leq 0$) and the same hypothesis for the submission phase (i.e., H_0: $\beta_2 + \beta_4 \leq \beta_3 + \beta_5$) to determine whether a salience bias exists among the contest winners. Our results suggest that both hypotheses are rejected at the 1% significance level. Hence, the salience bias is still persistent among winners in our case, thus affecting the outcomes of the contests, and Hypothesis 3 is supported.

Hypotheses 4A and 4B are two competing hypotheses. If the parallel path effect dominates in the learning phase, an increase in the number of contestants will reduce the effect of feedback on the score difference of the top contestants (i.e., $\alpha_4 + \alpha_5 < 0$). However, if the competition effect dominates, an increase in the number of contestants will amplify the effect of

Table 3. Ranking Correlations

Sample	Correlation	Sample	Correlation
Top 100 on public leaderboard	0.4516***	Top 100 on private leaderboard	0.5396***
Top 50 on public leaderboard	0.2543***	Top 50 on private leaderboard	0.3716***
Top 25 on public leaderboard	0.1249***	Top 25 on private leaderboard	0.2178***
Top 10 on public leaderboard	0.0367	Top 10 on private leaderboard	0.1378***
Top 5 on public leaderboard	−0.0103	Top 5 on private leaderboard	0.0732*
Public leaderboard winners	0.0367	Private leaderboard winners	0.0734

$^{*}p < 0.1$; $^{***}p < 0.01$.

feedback on the score difference of the top contestants (i.e., $\alpha_4 + \alpha_5 > 0$). In column (3) of Table 4, the sum of the last two interaction terms (i.e., $\alpha_4 + \alpha_5$) is positive at the 5% significance level, which supports Hypothesis 4B that the competition effect dominates the parallel path effect during the learning phase; therefore, as the number of contestants increases, the winners (and so the outcome of the contests) are more likely to be influenced by the salience bias during the learning phase.[12]

During the submission phase, the competition effect is at its minimum. As suggested by Hypothesis 5, the negative impact of the salience bias among the winners will be reduced when the number of contestants increases. In other words, the likelihood of a solution being selected as a final solution by the winners will be less associated with the public scores and/or more associated with the private scores as the number of contestants increases (i.e., $\beta_6 + \beta_8 < 0$ and/or $\beta_7 + \beta_9 > 0$). In column (3) of Table 5, the sum of the two interactions $\log(\#Team_c) \times \widetilde{RankInTeam}_{ijc}^{\text{Public}}$ and $Winner_{ic} \times \log(\#Team_c) \times \widetilde{RankInTeam}_{ijc}^{\text{Public}}$ (i.e., $\beta_6 + \beta_8$) is insignificant, and the sum of the two interactions $\log(\#Team_c) \times \widetilde{RankInTeam}_{ijc}^{\text{Private}}$ and $Winner_{ic} \times \log(\#Team_c) \times \widetilde{RankInTeam}_{ijc}^{\text{Private}}$ (i.e., $\beta_7 + \beta_9$) is positive at the 5% significance level. This suggests that among the winners, the solutions with high private scores are more likely to be selected as final solutions when there are more contestants. This suggests that the parallel path effect dominates during the submission phase. Hence, Hypothesis 5 is supported.[13]

Table 4. Results of the Learning Phase Model

	(1)	(2)	(3)
$Time_{ijc}$	−0.0007494	−0.0008435	0.0006821
	(0.009365)	(0.009352)	(0.009348)
$\log(\#Feedback_{ijc})$	0.05633***	0.05499***	0.05529***
	(0.003795)	(0.003809)	(0.003805)
$Winner_{ic} \times \log(\#Feedback_{ijc})$		0.04155	0.06762**
		(0.02686)	(0.03196)
$\log(\#Team_c) \times \log(\#Feedback_{ijc})$			−0.02062***
			(0.003250)
$Winner_{ic} \times \log(\#Team_c)$ $\times \log(\#Feedback_{ijc})$			0.07869***
			(0.02478)
Contestant fixed effect	Yes	Yes	Yes
Contest fixed effect	Yes	Yes	Yes
R^2	0.254	0.254	0.255
Observations	695,622	695,622	695,622

$^{**}p < 0.05$; $^{***}p < 0.01$.

6.4. Identification of the Salience Bias

The empirical approach used in this study is different from one assuming a typical causal relationship, as identified in econometrics analysis, in that we are not directly modeling the decision process of each team. Instead, we examine in retrospect whether the pattern we observe in the data is consistent with Pattern A (i.e., the pattern we should observe in the absence of salience bias) or Pattern B (i.e., the pattern we should observe in the presence of salience bias). By showing that the pattern we observe is consistent with Pattern B

Table 5. Results of the Submission Phase Model

	(1)	(2)	(3)
$Time_{ijc}$	1.3102***	1.3102***	1.3112***
	(0.04683)	(0.04685)	(0.04745)
$\widetilde{RankInTeam}_{ijc}^{\text{Public}}$	14.531***	14.600***	15.516***
	(0.2087)	(0.2090)	(0.2203)
$\widetilde{RankInTeam}_{ijc}^{\text{Private}}$	0.7217***	0.6571***	0.5345***
	(0.08929)	(0.08946)	(0.09277)
$Winner_{ic} \times \widetilde{RankInTeam}_{ijc}^{\text{Public}}$		−3.5554***	−2.6709
		(1.2698)	(2.3572)
$Winner_{ic} \times \widetilde{RankInTeam}_{ijc}^{\text{Private}}$		3.7230***	6.0402***
		(0.8209)	(1.1453)
$\log(\#Team_c) \times \widetilde{RankInTeam}_{ijc}^{\text{Public}}$			2.5658***
			(0.1626)
$\log(\#Team_c) \times \widetilde{RankInTeam}_{ijc}^{\text{Private}}$			−0.07634
			(0.08523)
$Winner_{ic} \times \log(\#Team_c)$ $\times \widetilde{RankInTeam}_{ijc}^{\text{Public}}$			−1.2979
			(1.3716)
$Winner_{ic} \times \log(\#Team_c)$ $\times \widetilde{RankInTeam}_{ijc}^{\text{Private}}$			1.9316**
			(0.7602)
Pseudo-R^2	0.658	0.658	0.663
Observations[a]	670,838	670,838	670,838

Note. All fixed effects are canceled out during the estimation of the conditional logit model.

[a]24,784 observations were dropped because there was no variability within the group.

$^{**}p < 0.05$; $^{***}p < 0.01$.

and by eliminating other factors that may cause Pattern B, we conclude that salience bias is present. Hence, a concern here is whether the pattern we observe is driven by other alternative explanations. In this section, we address this concern by (1) utilizing exogenous variations in the salience level of public score feedback to show that the pattern we observe gets stronger when the feedback is more salient, as an additional direct support to our salience bias explanation, and (2) ruling out alternative explanations. The results of all of the analyses in this section are in the online appendix.

6.4.1. Exogenous Variation—Additional Cues. Some Kaggle contests remind contestants of the difference between public and private scores or stress the importance of private scores in their contest descriptions, while others do not. For example, one Kaggle contest's description explicitly mentions, "The predictions on test data will remain hidden on the 'private leaderboard' until they are revealed at the end of the challenge. The final ranking and determination of the winners will be based on the 'private leaderboard' test data results." Information like this provides an additional cue to the contestants by reminding them of the importance of the private scores, thus likely counterweighing the salient influence of the public scores. The contestants in such contests are expected to pay more attention to cross-validation, thus having a smaller salience bias. We thus use a dummy variable, $Reminder_c$, to indicate whether contest c explicitly mentions the difference between the public scores and the private scores or stresses the importance of private scores. If the pattern we observe is indeed driven by the salience bias, then the contestants in these contests should be less subject to salience bias, so the effect of salience bias should be smaller in these contests. The results, shown in Tables A1a and A1b in the online appendix, are indeed consistent with the salience bias explanation. In Table A1a in the online appendix, the coefficient of $Reminder_c \times \log(\#Feedback_{ijc})$ is negative and significant, meaning that the effect of the salience bias is smaller if $Reminder_c = 1$ in the learning phase. In Table A1b in the online appendix, the coefficient of $Reminder_c \times \widetilde{RankInTeam}^{\text{Public}}_{ijc}$ is negative and significant, and that of $Reminder_c \times \widetilde{RankInTeam}^{\text{Private}}_{ijc}$ is positive and significant, meaning that the solutions with high public scores are less likely to be selected as final submissions if $Reminder_c = 1$.

6.4.2. Exogenous Variation—Number of Feedback Scores. Another variation we observe is the number of feedback scores each team receives. The more feedback that is received, the more salient the feedback information is. Thus, by showing that the effect of the salience bias increases as the feedback information becomes more salient (i.e., as more feedback is received), we can directly support the salience bias explanation that we propose. We have already used the number of feedback scores as the key variable in the learning phase to directly capture the salience level of feedback information (in Table 4). As in the learning phase, we can also examine how the salience level of feedback information moderates the effect of the salience bias by including the total number of feedback scores each team receives ($\#TotalFeedback_{ic}$) as a moderating variable in the submission phase (its main effect is absorbed by the contestant fixed effect). In Table A2 in the online appendix, we show that when the public score is more salient (i.e., $\#TotalFeedback_{ic}$ is higher), the effect of the salience bias gets stronger. Hence, the results are consistent with the salience bias explanation.

6.4.3. Alternative Explanations. There are some possible alternative explanations behind the patterns that we observe in our results. In this section, we perform additional analyses and present additional evidence from the results of a survey of actual Kaggle contestants to rule out these alternative explanations.

First, our results could be driven by the data given, for example, if the training set is smaller than or systematically different from the test set, or the portion of the test set used for deriving public scores is larger than the portion of the test set used for deriving private scores. To address this concern, we conduct a subsample analysis by excluding the contests that (1) mention any difference between the training set and the test set,[14] (2) have a training set smaller than the test set, or (3) use over 50% of the test set to calculate public scores. The results (Tables A3a and A3b in the online appendix) are consistent with our main results in both phases.

Second, contestants may not be aware of or may not be good at cross-validation. To rule out this alternative explanation, we conduct a subsample analysis by including only the contestants who have expressed their opinions about cross-validation on the forum for the contests they participated in. If a contestant has participated in discussions related to cross-validation, then it is reasonable to assume that the contestant knows what cross-validation is or learned what cross-validation is after these discussions occurred. The results are in Tables A4a and A4b in the online appendix. In column (1), we analyze the solutions by these contestants in all of the contests they participated in. In column (2), we analyze the solutions by these contestants only in the contests they participated in after the cross-validation discussion occurred. Both tables show that the results are consistent with the main results. We also examine the results for high-ability contestants, such as contest winners and experienced contestants who should know about cross-validation. The analyses are in Sections 6.5 and 6.6, and are consistent with our main results.

Finally, we conducted a survey of actual Kaggle contestants to get a better understanding of their behavior, which helps us establish the link between the observed contest results on Kaggle and the possible salience bias that might be induced by the feedback information. We sent a survey to 490 actual Kaggle contestants and received 101 responses.[15] A copy of the survey is provided in the online appendix. Based on the responses, we observe the following: (a) Almost all of the respondents (93 out of 101) know what *public score* and *private score* mean, and their understanding is consistent with the definitions provided on the Kaggle website. (b) Almost all of the respondents (99 out of 101) are aware of and have done cross-validation in the Kaggle contests that provide public scores. (c) Most of the respondents (77 out of 101) agree that cross-validation is as reliable as or more reliable than the public scores. (d) In retrospect, the majority of the respondents (68 out of 101) think that the solutions selected based on cross-validation have higher private scores than the solutions selected based on the public scores, and very few (13 out of 101) think that the solutions selected based on public scores have higher private scores. (e) In retrospect, most of the respondents (79 out of 101) think they should have paid more attention to cross-validation.[16]

These observations suggest that Kaggle contestants do receive cues from both public scores and cross-validation, they consider cross-validation more reliable, and, in retrospect, they wish they had done more cross-validation and selected solutions based on cross-validation. However, the patterns we observe in the data suggest that the contestants still overemphasize the feedback from public scores, rather than rely on their own cross-validation. This is likely due to the cognitive bias (i.e., the salience bias) that they were not aware of when they were competing in the contests. The survey results confirm that contestants are aware of alternative evaluation methods for their solutions (i.e., cross-validation) and acknowledge that they may be superior to the public scores provided by Kaggle. However, the data collected from Kaggle and our analysis suggest that contestants' behavior appears to be more consistent with the behavior driven by public scores.

6.5. Robustness Analyses

In this section, we run additional specifications of our model to ensure that our results are robust. First, every time a contestant uploads a solution online, the contestant receives the public score for the solution as the feedback. Because the number of uploaded solutions and the score difference may both be affected by the unobserved individual characteristics, in addition to including the contestant fixed effect, we model the number of feedback scores as an endogenous variable in our first robustness analysis. Specifically, we use the maximum number of solutions allowed ($MaxSolutionsAllowed_{ijc}$), which equals the number of solutions allowed per team per day multiplied by the number of days elapsed since the first solution was submitted by the team, as the instrumental variable (IV) for the number of feedback scores. Because the maximum number of solutions allowed is predetermined at the beginning of the contest and contestants tend to upload more solutions when they are allowed, this IV is strongly correlated with the number of feedback scores received but is uncorrelated with the error term, thus making it a valid IV for our analysis. Our results are shown in Table A5 in the online appendix. Because the number of feedback scores shows up only in the learning phase, we only do this analysis for the learning phase model. All of the results remain qualitatively unchanged.[17]

Second, instead of using a *Winner* dummy, we conduct a subsample analysis by using only the solutions of winners to directly test Hypotheses 3–5. Our results are in Tables A6a and A6b in the online appendix and remain qualitatively unchanged.

Third, because the number of participating teams changes over time during the contests, contestants may observe the number of contestants at different time points. Although this has a relatively small effect on the submission phase (because it typically happens at the end of a contest), this may change the influence of the competition effect at different times during the learning phase, depending on the number of participating teams at the moment the contestants are working on a new solution. Hence, instead of using the total number of teams in contest c ($\#Team_c$), we use the number of participating teams before contestant i uploaded solution j in contest c ($\#team_{ijc}$) to reexamine the competition effect during the learning phase. The results are in Table A7 in the online appendix and are consistent with our main results.

Fourth, to ensure that our results are not driven by excluding the contests with nonmonetary rewards, we include the four contests that provided only Kaggle points as rewards and rerun our analysis. The results are in Tables A8a and A8b in the online appendix and remain qualitatively the same.

Fifth, in the submission phase models, the high correlation between the public score and private score rankings may raise the concern of multicollinearity. Multicollinearity inflates the standard errors of the estimated coefficients, making it less likely for us to find significant results. Since our results are based on coefficients that are significant even with the inflated standard errors, we are confident that our results are robust despite possible multicollinearity. We also conduct an additional analysis in which we transform one of the independent variables to reduce their correlations. Specifically, we replace the private score ranking variable $\widetilde{RankInTeam}_{ijc}^{\text{Private}}$ with the difference between the public score and private score rankings, which is significantly less correlated with the public score ranking.[18] The results are in Table A9 in the online appendix and are consistent with our main results.

6.6. Individual Characteristics and Contest Characteristics

Given that a salience bias exists in crowdsourcing contests, an effective strategy for eliminating its effect should be sought. In this section, we explore how individual characteristics and contest characteristics potentially affect the influence of the salience bias on the Kaggle platform. In particular, experienced contestants tend to be more rational (List 2003), and therefore should be less subject to salience bias. For the contest characteristics, we investigate how the reward size moderates the influence of the salience bias. Previous studies have reported that extrinsic reward is a major incentive for contestants to invest effort in solving problems on such platforms (e.g., Liu et al. 2014). Hence, the size of a reward ($Rewardsize_c$) would have a moderating effect similar to that of a competition effect.

We take a log transformation of the variables $Experience_{ic}$ and $Rewardsize_c$ to correct the skewness of the variables (Gelman and Hill 2007). Then, we moderate the salience bias-related variables with these two characteristic variables. The results are in Tables A10a and A10b in the online appendix. The interaction $\log(Experience_{ic}) \times \log(\#Feedback_{ijc})$ in Table A10a is insignificant, suggesting that reliance on the public scores does not increase with experience during the learning phase. Table A10b further shows that, during the submission phase, the solutions with high public scores are less likely to be selected as final solutions by experienced contestants. Hence, contestants' experience may facilitate reducing salience bias.

The high statistical significance of the positive moderation effect in Table A10a suggests that rewards serve as a stimulus that drives the contestants to exert more effort, thus increasing the competition effect during the learning phase, but their effect in the submission phase is unsigned, consistent with the argument that the competition effect should be at the minimum in the submission phase.

7. Discussion and Conclusion

Elucidating the pitfalls and benefits of using platforms could assist future researchers and practitioners in integrating online platforms into their business strategies. This study uses a crowdsourcing platform to illustrate how behavioral biases can affect the productivity of the platform. Specifically, we examine how the negative impact of the salience bias can be eliminated or attenuated on a crowdsourcing platform. To our knowledge, this study is the first to address this issue, from which several crucial results can be obtained: First, a salience bias is persistent among all contestants (Hypotheses 1 and 2), including the winners, even when a parallel path effect is at work (Hypothesis 3). Second, the competition effect dominates the parallel path effect when the contestants exert more effort. Thus, the competition effect amplifies the influence of the salience bias on the winners during the learning phase (Hypothesis 4B). When the contestants need to exert less effort in a contest, the parallel path effect dominates, and this attenuates the influence of the salience bias on the winners during the submission phase (Hypothesis 5).

Studies related to crowdsourcing contests have typically followed the doctrine of rationality (e.g., Terwiesch and Xu 2008). However, without considering the behavioral aspects of the agents, the quality of the crowdsourcing outcome can be low. The findings of this study elucidate how individual biases can change the outcomes of crowdsourcing contests. Seekers and platform designers should be aware of the possibility of creating inferior outcomes because of systematic biases. One plausible remedy is to provide training to the contestants to raise their awareness of such behavioral anomalies.

Another implication of these results is that, although the competition effect fosters greater effort from high-ability contestants (Boudreau et al. 2016), if the contestants' objective is influenced by systematic bias, the competition effect may actually be detrimental to the overall outcome. For example, in our case, the high-ability contestants expend greater effort in the face of higher competition, but their effort is spent more on overfitting in response to the salient public scores. Thus, contest holders should be more aware of the downside of overincentivizing contestants.

In addition to Kaggle, other platforms (such as DrivenData, CrowdANALYTIX, and Datascience.net) also offer features similar to public scores. Therefore, our findings can be applied to other platforms that host predictive modeling contests. Although our results may not directly apply to other crowdsourcing sites that do not have a feature equivalent to public scores, our findings suggest that it is important for platforms and contest holders to be aware of the potential cognitive bias likely induced by in-progress feedback—especially when such feedback is not perfectly aligned with the objective of the contest. Furthermore, although this study investigates only the impact of the salience bias, the results potentially caution about the influence of other systematic biases as well as on crowdsourcing outcomes. For example, innovation processes tend to depend on leadership style and instruction (Basu and Green 1997, Ruscio and Amabile 1999). Seekers' perceptions of a certain type of solution or unintended misleading instruction might also systematically affect the outcome of the contestants' performance. Our study offers a cautionary note on the effects of design features and information availability on the overall performance of crowdsourcing workers.

Increasing the number of contestants can initiate both the parallel path effect and the competition effect. While the parallel path effect can attenuate the influence of salience bias on the winners' solutions, the competition effect works in the opposite direction. Contest holders should be aware of how contest characteristics such as the level of difficulty, the rewards, and the timing of the contest launch may affect the number of participants. Perhaps platform designers should also investigate the possibility of reducing systematic bias by manipulating the information presented on their platforms, for example, by making the information about the number of contestants less noticeable. Hence, the competition effect could be less severe and the parallel path effect could still be preserved because the overall number of contestants would remain the same. In addition, our results suggest that reminding the contestants about potential cognitive biases can help reduce their influence. For example, explicitly reminding contestants about the difference between the public and private scores in the context description or emphasizing the importance of cross-validation through data partitioning to look for generalized predictive models could remind contestants of the salience bias so that they can be less influenced.

We also investigate several contest- and individual-related characteristics that may affect the influence of the salience bias. The results provide additional suggestions for platform owners on how to reduce systematic bias and improve solution outcomes. In line with the explanation of related theories, when the reward size is larger, it also induces more effort in evaluating or creating solutions. Therefore, such contest features exacerbate the impact of the salience bias. Furthermore, we also find that experienced contestants may exhibit less cognitive bias and focus more on the final result. In contrast to conventional wisdom, where a crowd functions at its peak performance when it comprises a diverse group of workers (e.g., Ren et al. 2015), platform designers should encourage more experienced contestants to participate in their contests.

This study emphasizes a potential downside of providing in-progress feedback in crowdsourcing contests. Because we do not observe what would happen if public scores were not provided in our context, one of the limitations of this paper is that we are not able to draw conclusions on their overall effect, since providing in-progress feedback can also have upsides, such as encouraging participation. Another limitation is that although our empirical analyses, anecdotal evidence, and survey results suggest that salience bias exists on the platform, we are not able to directly capture the true impact of the salience bias. This is due to our inability to observe the cross-validation scores among contestants. These limitations should be addressed in future research.

Acknowledgments

The authors would like to thank the special issue editors, the associate editor, and the three anonymous reviewers for their thoughtful reviews and constructive suggestions during the review process.

Endnotes

[1] See https://www.ideaconnection.com/open-innovation-success/ (accessed June 20, 2017).

[2] For example, Sismanis (2010, p. 7), an experienced Kaggle contestant, shared his experience after he realized that overemphasizing the feedback while deemphasizing internal evaluations was a "bad" (irrational) choice: "[T]he discrepancies between my own cross-validations and the leader-board made me believe that much more complicated models were required, and I soon abandoned the basic principles of [devising] the winning submission. I tried more complicated models while relaxing the regularization efforts. In retrospect, this was a bad decision on my part; I should have realized earlier the importance of regularization, and the potential of overfitting for such a small test data set."

[3] See, for example, "Some General Thoughts on Overfitting," http://www.kaggle.com/general/12858 (accessed February 16, 2018).

[4] See "How to Select Your Final Models in a Kaggle Competition," http://www.chioka.in/how-to-select-your-final-models-in-a-kaggle-competitio/ and "The Dangers of Overfitting: A Kaggle Postmortem," http://gregpark.io/blog/Kaggle-Psychopathy-Postmortem/ (accessed February 16, 2018).

[5] See Coursera (https://www.coursera.org/specializations/machine-learning) and Udacity (https://www.udacity.com/course/intro-to-machine-learning–ud120; accessed June 18, 2017).

[6] We do not imply that the contestants exert no effort during the submission phase. (It is likely that they still exert some effort while selecting their final solutions.) For example, as indicated by Blohm et al. (2016), crowdsourcing workers exert cognitive effort to evaluate the quality of an idea. Instead, we argue that the parallel path effect is likely to dominate the competition effect because more of the effort has already been exerted during the learning phase.

[7] Our results remain qualitatively the same if we do not include the *Time* variable.

[8] The contestant and contest fixed effects (together with the other contestant or contest invariant variables) are canceled out in the conditional logit model.

[9] The private score of each solution is observed by the researchers, but it is not observed by the contestants during the contest.

[10] If the number of feedback scores was the same in two groups, we combined them into a single group.

[11] Our results are qualitatively the same if we cluster on both contests and contestants. The results are reported in Tables A11a and A11b in the online appendix.

[12] Because our goal is to examine the impact of salience bias on the outcomes of crowdsourcing contests, our focus is on the winners. It is interesting to note that the number of teams has an opposite effect on the nonwinners in Table 4. This can be explained by the overall effort-reducing effect suggested by Boudreau et al. (2011), which is different from the competition effect for high-ability contestants suggested by Boudreau et al. (2016). While increasing the number of

teams induces high-ability contestants to exert more effort to maintain their winning position (the competition effect), it discourages the low-ability contestants from exerting effort because their chances of winning will be further reduced (the overall effort-reducing effect). Therefore, the low-ability contestants work less aggressively in the direction guided by the salient feedback information, and thus are less subject to the influence of salience bias when there are more competitors in the learning phase (in Table 4).

[13] The number of teams has an opposite effect on the nonwinners in Table 5. This is because while having more teams increases the chance of getting a high-ability contestant to be the winner, it also increases the chance of getting more low-ability contestants participating. As a result, while the quality of the winning solution goes up because of the parallel path effect, the average quality of the nonwinning solutions may go down, because of the larger group of low-ability contestants who are more likely to be influenced by the salience bias.

[14] Our survey of actual Kaggle contestants suggests that most of the respondents (80 out of 101) consider the two sets from the same population if the contest does not explicitly mention how they split the data.

[15] We manually went over the list of contestants on Kaggle and collected 490 emails from their Kaggle profiles. A summary of the respondents' background information is given in Tables A12a–A12c in the online appendix.

[16] All of the results remain qualitatively the same if we focus on only a subsample of the survey respondents for which the average years of experience and the percentage of winners are comparable to those for the contestants in our data set.

[17] We also reran all of the other analyses in Sections 6.4–6.6 using the same IV for the number of feedback scores. All of the results remained qualitatively unchanged. The results are available on request. We thank an anonymous reviewer for this suggestion.

[18] We thank the associate editor for this suggestion.

References

Abu-Mostafa YS, Magdon-Ismail M, Lin HT (2012) *Learning from Data* (Self-published, AMLbook.com).

Adamczyk S, Haller J, Bullinger AC, Moslein KM (2011) Knowing is silver, listening is gold: On the importance and impact of feedback in IT-based innovation contests. Bernstein A, Schwabe G, eds., *Proc. 10th Internat. Conf. Wirtschaftsinformatik*, Vol. 1 (AIS, Atlanta), 221–230.

Alevy JE, Haigh MS, List JA (2007) Information cascades: Evidence from a field experiment with financial market professionals. *J. Finance* 62(1):151–180.

Ba S (2001) Establishing online trust through a community responsibility system. *Decision Support Systems* 31(3):323–336.

Ba S, Stallaert J, Whinston AB (2001) Research commentary: Introducing a third dimension in information systems design—The case for incentive alignment. *Inform. Systems Res.* 12(3):225–239.

Basu R, Green SG (1997) Leader-member exchange and transformational leadership: An empirical examination of innovative behaviors in leader-member dyads. *J. Appl. Soc. Psych.* 27(6): 477–499.

Bayus BL (2013) Crowdsourcing new product ideas over time: An analysis of the Dell IdeaStorm community. *Management Sci.* 59(1):226–244.

Becker BE, Huselid MA (1992) The incentive effects of tournament compensation systems. *Admin. Sci. Quart.* 37(2):336–350.

Berger J, Pope D (2011) Can losing lead to winning? *Management Sci.* 57(5):817–827.

Blohm I, Riedl C, Füller J, Leimeister JM (2016) Rate or trade? Identifying winning ideas in open idea sourcing. *Inform. Systems Res.* 27(1):27–48.

Bockstedt JC, Druehl C, Mishra A (2016) Heterogeneous submission behavior and its implications in innovation contests with public submissions. *Production Oper. Management* 25(7):1157–1176.

Boudreau KJ (2012) Let a thousand flowers bloom? An earlier look at large numbers of software app developers and patterns of innovation. *Organ. Sci.* 23(5):1409–1427.

Boudreau KJ, Lakhani KR (2013) Using the crowd as an innovation partner. *Harvard Bus. Rev.* 91(4):60–69.

Boudreau KJ, Lacetera N, Lakhani KR (2011) Incentives and problem uncertainty in innovation contests: An empirical analysis. *Management Sci.* 57(5):843–863.

Boudreau KJ, Helfat CE, Lakhani KR, Menietti M (2016) Performance responses to competition across skill-levels in rank order tournaments: Field evidence and implications for tournament design. *RAND J. Econom.* 47(1):140–165.

Burtch G, Ghose A, Wattal S (2014) Cultural differences and geography as determinants of online prosocial lending. *MIS Quart.* 38(3):773–794.

Butler B, Bateman P, Gray P, Diamant E (2014) An attraction-selection-attrition theory of online community size and resilience. *MIS Quart.* 38(3):699–728.

Camacho N, Donkers B, Stremersch S (2011) Predictably non-Bayesian: Quantifying salience effects in physician learning about drug quality. *Marketing Sci.* 30(2):305–320.

Connelly BL, Tihanyi L, Crook TR, Gangloff KA (2014) Tournament theory: Thirty years of contests and competitions. *J. Management* 40(1):16–47.

Dellarocas C (2003) The digitization of word of mouth: Promise and challenges of online feedback mechanisms. *Management Sci.* 49(10):1407–1424.

Dissanayake I, Zhang J, Gu B (2015) Task division for team success in crowdsourcing contests: Resource allocation and alignment effects. *J. Management Inform. Systems* 32(2):8–39.

Gelman A, Hill J (2007) *Data Analysis Using Regression and Multilevel/Hierarchical Models* (Cambridge University Press, Cambridge, MA).

Greene W (2008) *Econometric Analysis* (Prentice Hall, Upper Saddle River, NJ).

Griffin D, Tversky A (1992) The weighing of evidence and the determinants of confidence. *Cognitive Psych.* 24(3):411–435.

Henfridsson O, Bygstad B (2013) The generative mechanisms of digital infrastructure evolution. *MIS Quart.* 37(3):907–931.

Hosmer DW, Lemeshow S, Sturdivant RX (2013) *Applied Logistic Regression*, 3rd ed. (Wiley, Chichester, UK).

Huang Y, Singh PV, Srinivasan K (2014) Crowdsourcing new product ideas under consumer learning. *Management Sci.* 60(9): 2138–2159.

Jeppesen LB, Lakhani KR (2010) Marginality and problem solving effectiveness in broadcast search. *Organ. Sci.* 21(5):1016–1033.

Kahneman D, Tversky A (1972) Subjective probability: A judgment of representativeness. *Cognitive Psych.* 3(3):430–454.

Katz ML, Shapiro C (1985) Network externalities, competition, and compatibility. *Amer. Econom. Rev.* 75(3):424–440.

Lin M, Viswanathan S (2016) Home bias in online investments: An empirical study of an online crowdfunding market. *Management Sci.* 62(5):1393–1414.

List JA (2003) Does market experience eliminate market anomalies? *Quart. J. Econom.* 118(1):41–71.

List JA (2011) Does market experience eliminate market anomalies? The case of exogenous market experience. *Amer. Econom. Rev.* 101(3):313–317.

Liu TX, Yang J, Adamic LA, Chen Y (2014) Crowdsourcing with all-pay auctions: A field experiment on Taskcn. *Management Sci.* 60(8):2020–2037.

Parker GG, Van Alstyne MW (2005) Two-sided network effects: A theory of information product design. *Management Sci.* 51(10): 1494–1504.

Pope DG, Schweitzer ME (2011) Is Tiger Woods loss averse? Persistent bias in the face of experience, competition, and high stakes. *Amer. Econom. Rev.* 101(1):129–157.

Ren Y, Chen J, Riedl J (2015) The impact and evolution of group diversity in online open collaboration. *Management Sci.* 62(6):1–19.
Rochet J, Tirole J (2006) Two-sided markets: A progress report. *RAND J. Econom.* 37(3):645–667.
Rudi N, Drake D (2014) Observation bias: The impact of demand censoring on newsvendor level and adjustment behavior. *Management Sci.* 60(5):1334–1345.
Ruscio AM, Amabile TM (1999) Effects of instructional style on problem-solving creativity. *Creativity Res. J.* 12(4):251–266.
Shin SJ, Kim TY, Lee JY, Bian L (2012) Cognitive team diversity and individual team member creativity: A cross-level interaction. *Acad. Management J.* 55(1):197–212.
Sismanis Y (2010) How I won the "Chess Ratings–Elo vs. the Rest of the World" competition. Working paper, arXiv preprint arXiv:1012.4571.
Terwiesch C, Xu Y (2008) Innovation contests, open innovation, and multiagent problem solving. *Management Sci.* 54(9):1529–1543.
Tiefenbeck V, Goette L, Degen K, Tasic V, Fleisch E, Lalive R, Staake T (2018) Overcoming salience bias: How real-time feedback fosters resource conservation. *Management Sci.* 64(3):1458–1476.
Tilson D, Lyytinen K, Sørensen C (2010) Research commentary: Digital infrastructures: The missing IS research agenda. *Inform. Systems Res.* 21(4):748–759.
Tversky A, Kahneman D (1974) Judgment under uncertainty: Heuristics and biases. *Science* 185(4157):1124–1131.
Wooten JO, Ulrich KT (2017) Idea generation and the role of feedback: Evidence from field experiments with innovation tournaments. *Production Oper. Management* 26(1):80–99.
Yang Y, Chen PY, Pavlou P (2009) Open innovation: An empirical study of online contests. *Proc. 30th Internat. Conf. Inform. Systems* (AIS, Atlanta).
Yoo Y, Boland RJ, Lyytinen K, Majchrzak A (2012) Organizing for innovation in the digitized world. *Organ. Sci.* 23(5):1398–1408.