Public Leaderboard Feedback in Sampling Competition:

An Experimental Investigation*

Stanton Hudja† Brian Roberson‡ Yaroslav Rosokha§

September 2022

Abstract: We investigate the role of performance feedback, in the form of a public leaderboard, in a sequential-sampling contest with costly observations. We show theoretically that for contests with a fixed ending date (i.e., finite horizon), providing public performance feedback may result in fewer expected observations and a lower expected value of the winning observation. We conduct a controlled laboratory experiment to test the theoretical predictions, and find that the experimental results largely support the theory. In addition, we investigate how individual characteristics affect competitive sequential-sampling activity.

JEL classification: *D90, O31, C72, C90, D83*

*This paper also benefited from comments by participants at the 2018 Southern Economics Association Annual Meeting, 2019 American Economic Association Annual Meeting, 2019 Purdue-IU Experimental and Behavioral Mini Conference and seminar participants at Purdue University.

†Department of Economics, University of Toronto • Email: stanton.hudja@utoronto.ca

‡Krannert School of Management, Purdue University • Email: brobers@purdue.edu

§Krannert School of Management, Purdue University • Email: yrosokha@purdue.edu

1 Introduction

Public leaderboards are a common feature on popular innovation-contest platforms such as Kaggle.com, drivendata.org, and challenge.gov. The purpose of this study is to theoretically and experimentally examine whether the presence of a leaderboard necessarily increases contestant effort and the expected quality of the winning innovation. In particular, our framework for addressing this question builds upon the classic statistical problem of sequential sampling[1] in which a decision-maker, using the information from previous observations, chooses whether to make an additional costly observation.[2] The sequential sampling problem's emphasis on the trade-off between continued exploration and stopping to exploit the value of a given sequential random sample is clearly relevant for a wide variety of applications.[3] In this context, we experimentally test parameter configurations where theory predicts that a public leaderboard results in decreased levels of contestant effort and expected quality of the winning innovation and find that the presence of a public leaderboard does not necessarily improve innovation-contest outcomes.

To understand how leaderboard feedback affects the competition, note that the presence

[1]This problem appears to have been first formulated in Wald (1947). Early work includes Robbins (1952), Bradt and Karlin (1956), Feldman (1962), and Berry (1972). In economics, early applications include Stigler (1961) and the following literature on search, and Rothschild (1974) and the following literature on two-armed bandits.

[2]In sequential sampling, each observation, or draw, of an innovation quality may be thought of as either a new innovation or a quality improvement to a previous innovation.

[3]For an introduction to sequential-sampling problems, see DeGroot (1970). In addition to innovation competition, which we discuss in more detail below, recent applications include, among others, dynamic public-goods problems (Keller, Rady and Cripps, 2005), long-term contracts (Halac, Kartik and Liu, 2016), moral hazard in teams (Bonatti and Hörner, 2011), voting for reforms (Strulovici, 2010; Khromenkova, 2015), and decision timing (Fudenberg, Strack and Strzalecki, 2018).

of a leaderboard generates two distinct effects on the dynamics of effort provision that are not present with private feedback. With a leaderboard, the trailing competitor (henceforth, *follower*) may condition her choice of whether or not to make an additional costly observation on the leader's score. When the leader's score is low, the follower is more likely to be able to overtake the leader, and thus, leaderboard feedback may encourage followers to continue searching. Conversely, when the leader's score is high, a follower is less likely to be able to overtake the leader, and thus, leaderboard feedback may discourage followers from continuing to search. We show that in equilibrium: (i) followers who trail in the competition are more likely to invest in additional search than leaders, and (ii) all competitors reduce their search efforts as the leader's existing innovation quality increases. The results of our experiment confirm these theoretical predictions that current leaders tend to exert less search effort than followers and that both leaders and followers become less willing to exert search effort as the leader's innovation quality increases.

The dynamics of innovation effort provision help provide insight as to why leaderboard feedback may result in a lower expected value of the winning innovation for contests with a fixed ending date. In particular, a fixed ending date presents an obstacle for a follower that is attempting to overtake the leader. As a result, the leader score at which the leaderboard starts to discourage follower effort decreases as the length of the contest decreases. Because contest length has a less pronounced discouragement effect on the private-feedback contest, we find that there exist fixed contest lengths that, given the other model parameters, are sufficiently short as to result in leaderboard feedback generating a lower equilibrium expected quality for the winning innovation than the corresponding private-feedback contest.

An additional consideration with leaderboard feedback is its potential to generate an escalation of commitment (i.e., sunk-cost fallacy) that is reminiscent of the dollar auction and the penny auction.[4] That is, with a leaderboard, the follower knows that he or she is not in the

[4]See, for example, Hinnosaar (2016) on the penny auction and Shubik (1971) and O'Neill (1986) on the dollar auction.

lead and may consider his or her sunk research costs when deciding whether to try to take the lead by making an incremental investment in additional research effort. We investigate how individual characteristics, including sunk-cost fallacy, affect competitive sequential-sampling activity. In the experiments, we find that performance on a sunk-cost-fallacy elicitation task is a significant predictor of behavior both with and without leaderboard feedback when there are four competitors in the contest, but not when there are two competitors in the contest. In addition, we find that risk aversion is a significant predictor of behavior both with and without leaderboard feedback regardless of the number of competitors. Importantly, we find that the direction of these effects is consistent with our theoretical predictions.

Our paper contributes to several active streams of literature. First, we contribute to the experimental literature on feedback in innovation contests. There are several recent examples of experimental work that examine potential drawbacks of providing feedback in related contest environments, including Kuhnen and Tymula (2012), Ludwig and Lünser (2012), and Deck and Kimbrough (2017). Most closely related is Deck and Kimbrough (2017) who experimentally examine the exponential-bandit based innovation competition in Halac, Kartik and Liu (2017).[5] In that setting, Deck and Kimbrough (2017) find that withholding information leads to better innovation outcomes. This result arises from the fact that the information that your opponents have not procured the zero-one innovation lowers your own belief about the probability that innovation is possible. That is, information may only be discouraging, and thus, hiding information may be valuable. In a variation of a two-stage difference-form contest, Ludwig and Lünser (2012) find that feedback influences the dynamics of effort provision but not total effort. Kuhnen and Tymula (2012) find a similar result in an experiment that is modeled as a single-stage difference-form contest that is repeatedly played and feedback affects ego utility which may evolve over time. Lastly, in a recent survey, Dechenaux, Kovenock and Sheremeta (2015) highlight that in environments

[5]Recent bandit experiments also include, Rosokha and Younge (2020), Hoelzemann and Klein (2021), Hudja (2021), and Banovetz and Oprea (2022).

where it is difficult for the follower to overtake the leader, feedback may result in the trailing player dropping out (e.g., Fershtman and Gneezy, 2011) or reducing their effort (e.g., Malueg and Yates, 2010). In the case of sequential-sampling competition, our experimental results are consistent with some of the findings on the dynamics of effort provision observed in these papers. In particular, we find that followers who trail in the competition are more likely to continue to search than leaders, and all competitors reduce their search effort as the leader's existing innovation quality increases and it becomes more difficult for the follower to overtake the leader.

Second, our work is related to the literature on factors that motivate individuals to innovate. In particular, on the experimental side, recent studies have examined the role of incentives (Ederer and Manso, 2013), preferences (Herz, Schunk and Zehnder, 2014; Rosokha and Younge, 2020), and biases (Herz, Schunk and Zehnder, 2014). On the empirical side, two recent surveys by Astebro et al. (2014) and Koudstaal, Sloof and Van Praag (2015) highlight that entrepreneurs are typically less risk and loss averse. In the current paper, we consider the extent to which risk aversion, loss aversion, and the sunk-cost fallacy play a role in sequential-sampling competition.[6] Specifically, as part of our experiment, we elicited those three measures with incentivized multiple-price list tasks. We find that risk aversion is a significant predictor of the number of costly innovation actions in the contest, with more risk-averse subjects taking fewer actions. We also find that sunk-cost fallacy matters in larger contests. Specifically, subjects who exhibit sunk-cost fallacy in the elicitation task take more innovation actions when they are followers in the contest. At the same time, we did not find that our measure of loss aversion was predictive of subjects' behavior.

[6]We focus on risk aversion and loss aversion as characteristics that have been documented to matter in the lab (e.g., Herz, Schunk and Zehnder, 2014; Rosokha and Younge, 2020) and field (Astebro et al., 2014; Koudstaal, Sloof and Van Praag, 2015) settings. In addition, we consider the sunk-cost fallacy because it has been shown to affect behavior in a related setting of penny auctions (Augenblick, 2015).

Finally, we contribute to the literature on innovation competition. Existing approaches include but are not limited to variations of all-pay auctions (e.g., Che and Gale, 2003; Chawla, Hartline and Sivan, 2015), exponential-bandit contests (e.g., Halac, Kartik and Liu, 2017; Bimpikis, Ehsani and Mostagir, 2019), two-stage difference-form contests (e.g., Aoyagi, 2010; Klein and Schmutzler, 2017; Goltsman and Mukherjee, 2011; Gershkov and Perry, 2009; Mihm and Schlapp, 2018; Yildirim, 2005), crowdsourcing contests (e.g,. Terwiesch and Xu, 2008; DiPalantino and Vojnovic, 2009; Erat and Krishnan, 2012; Ales, Cho and Körpeoğlu, 2017), dynamic contests (e.g., Lang, Seel and Strack, 2014; Seel and Strack, 2016), and structural/empirical models of innovation contests (e.g. Gross (2017); Lemus and Marshall (2021)). For example, Lemus and Marshall (2021) examine Markov Perfect equilibrium in a variation of continuous-time sequential-sampling competition which includes features such as: (i) new contestants exogenously entering the competition at a constant rate over time and (ii) for each contestant innovation opportunities arrive stochastically over time. In this framework, Lemus and Marshall (2021) find that the effect of leaderboard feedback is theoretically ambiguous.[7]

In contrast, our work is most closely related to classic sequential-sampling competition, as in Taylor (1995), Fullerton and McAfee (1999), Baye and Hoppe (2003), and Rieck (2010), which readily lends itself to both multi-period competition and standard exploration versus exploitation considerations. Within this line of research, Fullerton and McAfee (1999) and Baye and Hoppe (2003) consider the case of no feedback and Taylor (1995) considers the case of private feedback. Our focus in this study is on leaderboard feedback in a setting with an arbitrary, but fixed, number of periods and in which the contestants may have general utility

[7]Lemus and Marshall (2021) estimate their model on data obtained from kaggle.com for competitions with a public leaderboard. The authors then run a series of counterfactual simulations to show a positive effect of leaderboard feedback on the number of submissions and the quality of winning submission. The authors also conduct a set of student competitions on kaggle.com to experimentally support their results.

functions. The closely related/special case of our model with two periods, two risk-neutral players and observable actions is examined in Rieck (2010), who find that private feedback generates a higher equilibrium expected value of the winning innovation. Conversely, we find that with a sufficiently long horizon, leaderboard feedback results in higher equilibrium expected values for the winning innovation than private feedback does. However, there exists a range of finite contest lengths that are sufficiently short that leaderboard feedback generates lower equilibrium levels of contestant effort and expected values for the winning innovation than the corresponding private-feedback contest.

The rest of the paper is organized as follows: in section 2, we present the theoretical model. In section 3, we provide details of the experimental design. In section 4, we develop predictions for our environment and organize them into four hypotheses. In section 5, we present the main results of the experiment. In section 6, we explore the robustness of the results to increasing the number contest participants. Finally, in section 7, we conclude.

2 Theory

Consider an N-player T-period dynamic innovation contest, along the lines of Taylor (1995). In this model, innovation activity takes the form of a search process with perfect recall. In each period $t \in \{1, \ldots, T\}$, each player $i \in \{1, \ldots, N\}$ has the opportunity to exert effort at a cost of $c > 0$. If player i exerts effort in period t, then at the beginning of period $t+1$ she obtains an innovation with quality level $s_{i,t+1}$, a random variable that is distributed according to F where F has a continuous and strictly-positive density everywhere on its support, which is assumed to be a convex subset of $\mathbb{R}_+$ with a lower bound of 0.[8] In the event that player i does not exert effort in period t, let $s_{i,t+1} = 0$. Player i's period t choice of effort is represented as the period t action $a_{i,t} \in \{D, ND\}$, where D denotes that effort is exerted and a draw is made from F and ND denotes that effort was not exerted and no draw was made from F. Player i's innovation "score" at the beginning of period

[8]In the experiment, we assume that innovations are exponentially distributed ($F(x; \lambda) = 1 - e^{-\lambda x}$ and $f(x; \lambda) = \lambda e^{-\lambda x}$, where $\lambda > 0$ is the rate parameter).

$t \geq 2$ is denoted by $\overline{s}_{i,t} \equiv \max\{s_{i,2}, \ldots, s_{i,t}\}$. After T periods, the contest ends and the player with the highest innovation score at the end of period T, that is, the player i with $\overline{s}_{i,T+1} = \max\{\overline{s}_{1,T+1}, \ldots, \overline{s}_{N,T+1}\}$, is awarded a prize with value $v \geq Nc$.[9] In the case of a tie, the winner is randomly chosen.

We examine two levels of feedback in the dynamic-innovation contest: (i) private feedback and (ii) leaderboard feedback. In the private-feedback innovation contest, each player i knows her current score ($\overline{s}_{i,t}$) at the beginning of each period t. In the leaderboard-feedback innovation contest, each player i knows, in addition to her own private feedback, the current max score,[10] $\max\{\overline{s}_{1,t}, \overline{s}_{2,t}\}$ at the beginning of each period t. Note that the leaderboard-feedback innovation contest is a game with imperfect public monitoring.[11] That is, players do not directly observe the actions of other players but only have noisy (public) information about action profiles in previous periods.

To illustrate how noisy (public) information in the leaderboard-feedback innovation contest affects the players' information sets, consider an arbitrary period in which the current max score is strictly positive. Because the players have private information regarding their own actions in previous periods, no player is able to deduce which of the other players chose to draw (D) in the previous period. Clearly, the nature of the information sets in this game rules out the use of a backward induction process based on proper subgames. However, utilizing the standard approach to imperfect public monitoring games, we can recover a concept of sequential rationality by focusing on a restriction of the strategy space. In the

[9]For the remaining cases of $v \in [0, Nc)$, note that if $c > v$ then the contest is trivial, and it is straightforward to extend our analysis to the case of $v \in [c, Nc)$.

[10]Note that the characterization of the perfect public equilibrium of the leaderboard-feedback game in Section 2.1, also applies to the variation of the game in which at the beginning of each period, each player observes all players' scores.

[11]For more information on games with imperfect public monitoring, see Mailath and Samuelson (2006).

following subsection, we characterize the perfect public equilibrium (henceforth PPE) for the leaderboard-feedback innovation contest.

Before proceeding, note that we use the convention, due to Taylor (1995), of referring to each draw of an innovation quality $s_{i,t}$ as a new innovation. Recall that an equivalent interpretation is that player i is working on one specific innovation and that each draw of an innovation quality $s_{i,t}$ is in regards to searching over quality improvements to that particular innovation. Depending on the application, this second interpretation may be more natural.

2.1 PPE in Leaderboard-Feedback Innovation Contests

In Online Appendix A, we characterize the PPE in the leaderboard-feedback innovation contest for the case of a general utility function that allows for risk aversion, and in Online Appendix B, we address the modeling of loss aversion and sunk-cost fallacy considerations. For expositional purposes, we focus here on the case of $N = 2$ risk-neutral players and relegate the details of the general N-player case to Online Appendix A.[12] Before presenting our results on the leaderboard-feedback innovation contest, we very briefly review the results, due to Taylor (1995), for the private-feedback innovation contest.

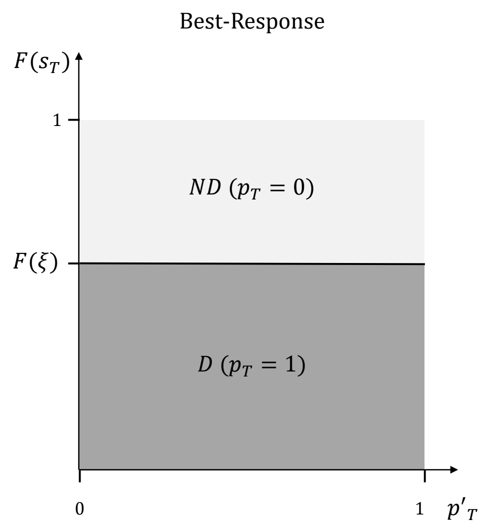
Private Feedback

The private-feedback innovation contest is examined in Taylor (1995). In particular, Proposition 2 of that paper establishes that equilibrium takes the form of a stopping rule in which each player i continues to exert effort until her max score hits a threshold – denoted by ξ – and then she stops exerting effort. Given that the other player is using a stopping rule with threshold ξ, Figure (1) presents player i's best response correspondence as a function of player i's private score at the beginning of period T, $s_{i,T}$, and the probability that the opponent draws in period T, which is denoted by $p_{-i,T}$. Note that because player i is using a stopping rule with threshold ξ, player $-i$ draws (i.e. $p_{-i,T} = 1$) if player $-i$'s private score

[12]Note that for a closely related version of the leaderboard-feedback innovation contest with $N = 2$ risk-neutral players, $T = 2$ periods, and observable actions, Rieck examines the corresponding set of subgame perfect equilibrium.

at the beginning of period T, $s_{-i,T}$, is less than ξ. Otherwise, player $-i$ does not draw (i.e. $p_{-i,T} = 0$). From Figure (1) we see that player i's best response to player $-i$ using a stopping rule with threshold ξ is to use a stopping rule with threshold ξ.

Figure 1: **Period T Local Best Response for Private Feedback**



Notes: s_T – own score in period T; $F(.)$ – distribution of innovation quality; p'_T – probability that the other player draws in period T; $ND(p_T = 0)$ – decision not to draw; $D(p_T = 1)$ – decision to draw; ξ – threshold determined by equation (1).

The equilibrium value of the threshold ξ is determined by the equation

$$v \int_{\xi}^{\infty} (1 - F^T(\xi)) \frac{F(x) - F(\xi)}{1 - F(\xi)} dF(x) - c = 0. \tag{1}$$

For example, in our experiment, we assume that when a player exerts effort in a given period, the quality of the innovation in that period is a random variable that is distributed according to $F(x; \lambda) = 1 - e^{-\lambda x}$ with $\lambda = 0.125$, which implies that for $T = 10$, the equilibrium stopping rule has a threshold of $\xi = 12.16$.

Leaderboard Feedback

Let f_t (l_t) denote the follower (leader) in an arbitrary period t. To recover a concept of sequential rationality in this imperfect public monitoring game, we restrict our focus to public strategies, where a strategy is public if in every period t the strategy depends only on the public history. Given the definition of public strategies, we define a PPE as follows. A PPE is a profile of public strategies that specifies a Nash equilibrium for each public history. To characterize the set of PPE, we use a recursive procedure that begins with the longest possible public histories (i.e. the beginning of period T) and moves back through the game tree.

Given a (public) leader score of s_T at the beginning of the final stage T, note that the probability that a stage T draw by the follower does [does not] overtake a leader who does not draw in stage T is $1 - F(s_T)$ [$F(s_T)$]. Similarly, the probability that a stage T draw by the follower does [does not] overtake a leader who also draws in stage T is $F(s_T)(1 - F(s_T)) + \frac{(1-F(s_T))^2}{2} = {(1-(F(s_T))^2)}/{2}$ $\left[{(1+(F(s_T))^2)}/{2}\right]$. In the final period T, if the max score at the beginning of period T is s_T, then the continuation game corresponds to the following matrix game:

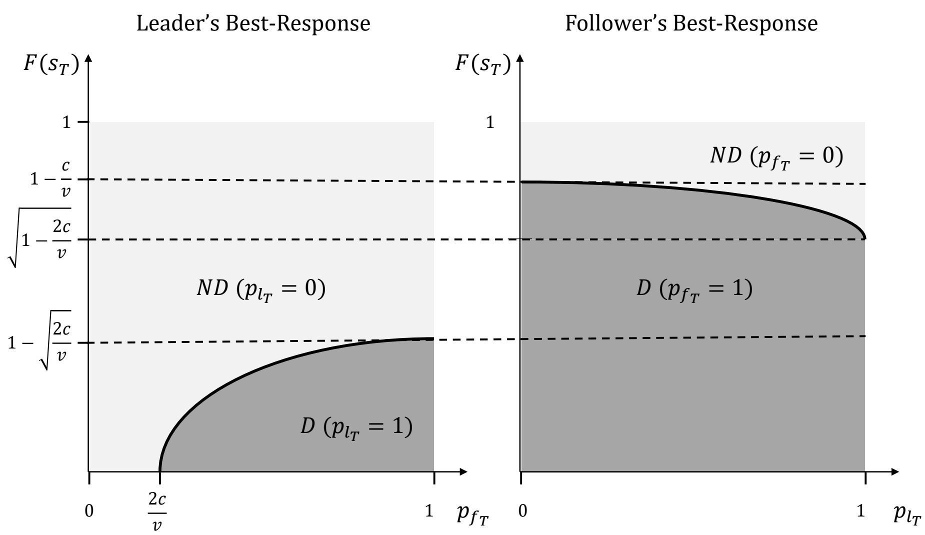
Table 1: **Period T Continuation Game for (Public) Leader Score s_T**

Leader (l_T) \ Follower (f_T)	D	ND
D	$\frac{v\left(1+(F(s_T))^2\right)}{2} - c$, $\frac{v\left(1-(F(s_T))^2\right)}{2} - c$	$v - c$, 0
ND	$vF(s_T)$, $v\left(1 - F(s_T)\right) - c$	v , 0

From Table 1, we see that the period T follower's (f_T's) final-stage local expected payoff from choosing to draw (D) when the period T leader (l_T) chooses not to draw (ND) is $v(1 - F(s_T)) - c$. Similarly, f_T's expected payoff from choosing D when l_T chooses D is $\frac{v(1-F(s_T)^2)}{2} - c$. Regardless of l_T's period T action, the payoff to f_T from choosing ND in period T is 0. The expected payoffs for the period T leader (l_T) follow along similar lines.

To calculate the final-stage local equilibrium, let p_{l_T} (p_{f_T}) denote the probability that the period T leader l_T (period T follower f_T) draws in period T. Figure 2 presents the players' best-response correspondences as a function of the leader's max score at the beginning of period-T, s_T, and of the probability that the opponent draws in period T and receives a stochastic period-T innovation quality distributed according to $F(\cdot)$.

Figure 2: **Period T Local Best Responses for Leaderboard Feedback**



Notes: s_T – score in period T; $F(.)$ – distribution of innovation quality; p_{f_T} – probability that follower draws in period T; p_{l_T} – probability that the leader draws in period T; $ND(p_{i_T} = 0)$ – decision not to draw by player $i \in \{leader, follower\}$; $D(p_{i_T} = 1)$ – decision to draw by player $i \in \{leader, follower\}$;

Proposition 1 characterizes the final-stage local equilibrium strategies and expected payoffs that follow directly from the best-response correspondences given in Figure 2. In particular, if $F(s_T) \in \left[0, 1 - \sqrt{\frac{2c}{v}}\right]$ and $p_{f_t} = 1$, then we see from the Leader's Best-Response panel of Figure 2 that $D(p_{l_T} = 1)$ is a best response for the the leader. Similarly, if $F(s_T) \in \left[0, 1 - \sqrt{\frac{2c}{v}}\right]$, then we see from the Follower's Best-Response panel of Figure 2 that for any value of $p_{f_t} \in [0, 1]$, the follower's best response is $D(p_{f_T} = 1)$. The remaining cases of $F(s_T) \in \left(1 - \sqrt{\frac{2c}{v}}, 1 - \frac{c}{v}\right]$ and $F(s_T) \in \left(1 - \frac{c}{v}, 1\right]$ follow along similar lines.

Proposition 1 *For any public history with period T leader score s_T, the final-stage local PPE strategies are characterized as follows:*

$$\begin{cases} \text{Both draw} & \text{if } F(s_T) \in \left[0, 1 - \sqrt{\frac{2c}{v}}\right] \\ \text{only follower draws} & \text{if } F(s_T) \in \left(1 - \sqrt{\frac{2c}{v}}, 1 - \frac{c}{v}\right] \\ \text{neither draws} & \text{if } F(s_T) \in \left(1 - \frac{c}{v}, 1\right] \end{cases}.$$

The corresponding final-stage local PPE expected payoffs for the leader and follower are given in Table 1.

Regarding intuition for the final-stage local equilibrium strategies, recall that as s_T increases, the probability of an additional draw overtaking the leader score decreases. When the leader score s_T is low enough that $F(s_T) \in \left[0, 1 - \sqrt{\frac{2c}{v}}\right]$, the probability of an additional draw overtaking s_T is sufficiently high that both the leader and the follower have incentive to invest in an additional innovation draw. For the intermediate values of s_T in which $F(s_T) \in \left(1 - \sqrt{\frac{2c}{v}}, 1 - \frac{c}{v}\right]$, the probability of an additional draw overtaking s_T is high enough that the follower has incentive to invest in an additional innovation draw but not so high that the leader also has incentive to draw. Lastly, for the remaining high values of s_T in which $F(s_T) \in \left(1 - \frac{c}{v}, 1\right]$ the probability of an additional draw overtaking s_T is sufficiently low that neither the leader nor the follower have incentive to invest in an additional innovation draw.

To calculate the (closed-form) PPE strategies, we may take the Proposition 1 final-stage local expected payoffs and work back through the game tree to stage $T - 1$. The only issue in continuing this recursive process all the way to the root of the game in stage 1 is the calculation of the expected continuation payoffs in the period t local continuation games. We provide details on these calculations in Online Appendix A.

3 Experimental Design

In this section, we describe the experimental design. In particular, the primary goal of the experiment is to address the role of feedback in sequential-search innovation competition. To this end, the main part of our experiment consists of two within-subject treatments: (i) a private-feedback treatment and (ii) a leaderboard-feedback treatment. In addition to the primary goal, our aim is to better understand factors that may influence individuals to innovate. To this end, after the two main treatments, our design includes an individual search task that removes the strategic aspect present in the two competitions and the elicitation of individual (e.g., risk aversion) and personality (e.g., grit) characteristics that may be important in an innovation setting. Next, we elaborate on details of the design and our implementation of the experiment.

3.1 Private-Feedback and Leaderboard-Feedback Contests

At the beginning of the experiment, each subject individually reads instructions that are displayed on their computer screen. In particular, we implemented a within-subject design, whereby each subject starts the experiment with either eight private-feedback contests or eight leaderboard-feedback contests and then switches to the other feedback type for contests 9 through 16. Thus, before contests 1 and 9, subjects are provided with detailed instructions and practice tasks that explain the setting of the upcoming eight contests. During the practice tasks, subjects were matched with a computer that made decisions randomly, and subjects were informed about the random behavior of the opponent in the practice task. All instructions used in the experiment are provided in Online Appendix C.

Each contest consists of two subjects matched for 10 periods of decision-making. Prior to the first period, each subject is given an endowment of \$10.00. Within each period, subjects have the opportunity to pay a cost $c = \$1.00$ to draw an innovation quality from an exponential distribution with parameter $\lambda = 0.125$. At the end of 10 periods, the contest ends and the subject with the highest-quality innovation (the highest score) wins the prize of $v = \$10.00$. Each subject keeps any money left over from her endowment. We chose these parameters because they provide interesting qualitative model predictions in a simple environment and were the same for the private and leaderboard treatments as well as for the individual search task described in section 3.2.

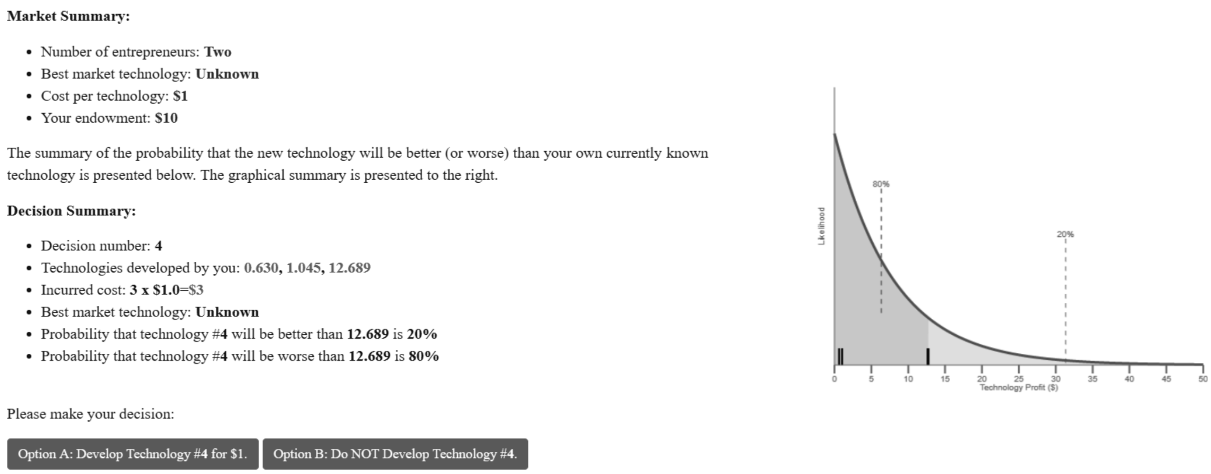
The first treatment is a two-player private-feedback contest in which each subject only receives feedback on their own innovations. Specifically, in each period, subjects decide whether to innovate. Although subjects know the quality of their own innovation, they do not know whether they are winning or losing until all decision periods are over. That is, the winning innovation is revealed only at the end of the contest. A screenshot of the private-feedback treatment is presented in Figure 3(a). In particular, during each period, each subject has access to the number of times she has drawn, the quality of each of the past innovations she has drawn, and her current innovation score (her innovation with the highest quality). To simplify decision-making, subjects are told the probability that an additional draw will result in a higher individual innovation score. At the end of the contest, subjects are informed of the winner of the contest and the amount of money they have earned for the contest.

The second treatment is a two-player leaderboard-feedback contest in which each subject receives feedback on her own innovation as well as the innovation that is currently leading the contest. Specifically, similar to the private-feedback contest, in each period of the leaderboard-feedback contest, subjects decide whether to innovate; however, the contest's best innovation is now revealed at the start of each period. Thus, each participant knows whether she is a leader or a follower. A screenshot of the leaderboard-feedback treatment is presented in Figure 3(b). Although most aspects of the leaderboard-feedback treatment are the same as in the private-feedback treatment, subjects receive additional feedback regarding the current highest score in the contest. That is, subjects always know whether they are currently winning or losing the contest and the probability that their next draw will result in their score being higher than the current maximum score.[13]

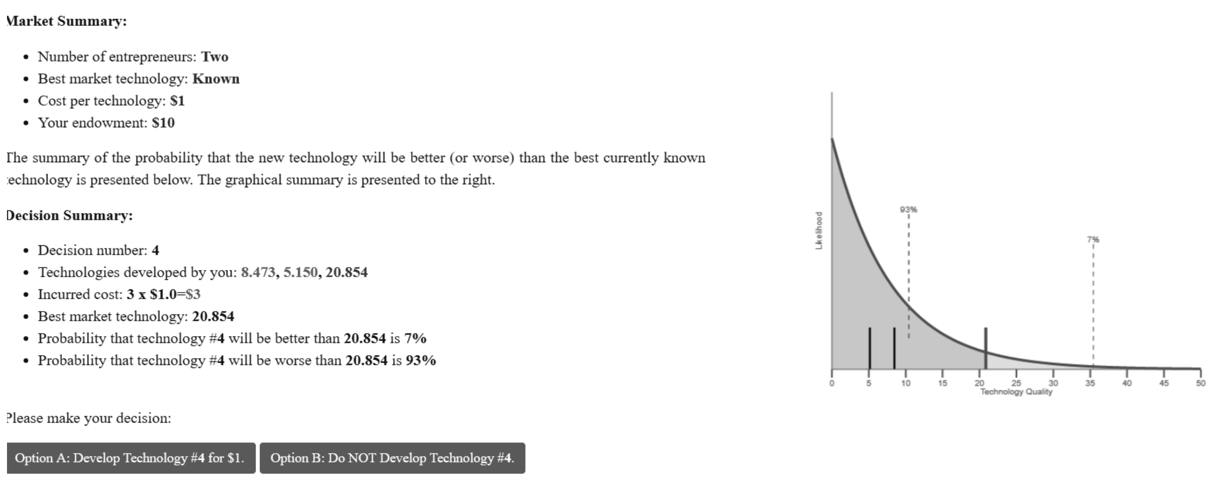
[13]Subjects are no longer shown the probability that an additional draw will result in a higher individual innovation score.

Figure 3: **Screenshots of the Experimental Interface**

(a) Private-feedback Treatment Screenshot



(b) Leaderboard-feedback Treatment Screenshot



3.2 Individual Tasks and Questionnaires

After completing both treatments, subjects were presented with several individual tasks. In particular, subjects completed three elicitation tasks: (i) a risk-aversion task, (ii) a loss-aversion task, and (iii) a sunk-cost-fallacy task. In each of these three tasks, subjects chose one of two options for each of the 20 decisions. The decisions were organized into a multiple price list as is common in the literature (e.g., Holt and Laury, 2002; Rubin, Samek and Sheremeta, 2018). In particular, the first task was the risk-aversion task. In this task, each participant chose between a risky option (50% chance of \$10.00 and a 50% percent chance of \$0.00) and a safe option that was varied across decisions (started at \$0.50 and increased by \$0.50 in each subsequent decision). The second task was the loss-aversion task. In this task, each participant chose between a safe option of \$0.00 and a risky option that had a 50% chance at \$5.00 and a 50% chance of a loss (varied from −\$0.50 to −\$10.00 in increments of \$0.50). The third elicitation task was the sunk-cost-fallacy task. In this task, subjects were given an endowment of \$15.00 and were required to pay \$5.00 to initiate a project in stage 1. Each subject then decided whether to complete the project at various completion costs in stage 2. Completing the project was always worth \$7.50; however, the cost varied between decisions. The completion cost started at \$0.50 and increased by \$0.50 in each subsequent decision. The sunk-cost fallacy occurs if the subject completes the project at a completion cost greater than \$7.50 in stage 2. Completing the project at a cost greater than \$7.50 is an over-investment, or escalation of commitment, that arises from previously sinking the \$5.00 into initiating the project. Screenshots of the three individual elicitation tasks are presented in Figures D1–D3 in the Online Appendix.

In addition to the above elicitation tasks, each subject participated in eight individual search tasks. The individual search tasks were similar to the two contests except that the human opponent was replaced with an existing innovation of a known quality. In particular, the existing innovation took on five values: 15.177, 16.832, 18.421, 20.205, and 23.966.[14] Each subject saw all five values, and the values 15.177, 18.421, and 23.966 were repeated

[14]These values correspond to the 85th, 88th, 90th, 92nd, and 95th percentiles of the exponential distribution, respectively. In particular, the risk-neutral agent would be indifferent between drawing and not drawing if the existing innovation was 18.421.

twice. The five values were displayed in random order. If the subject ended the period with an innovation of greater quality than the existing innovation, she won \$10.00. Thus, these tasks allow us to analyze individual behavior in a similar environment but without competition against another human subject. A screenshot of the individual search task is presented in Figure D4 in the Online Appendix.

The experiment concluded with three unincentivized personality questionnaires. In particular, the first questionnaire measured the psychological construct of grit through the 12-item Grit Scale (Duckworth et al., 2007). The second questionnaire measured the big five characteristics (agreeableness, extraversion, neuroticism, openness, and conscientiousness) through the 44-item big-five inventory (John and Srivastava, 1999). The third questionnaire measured achievement-striving and competitiveness through the 10- and 6-item scales obtained from the International Personality Item Pool.[15]

3.3 Experimental Administration

All parts of the experiment, including instructions, innovation contests, individual elicitation tasks, and personality questionnaires, were implemented in oTree (Chen, Schonger and Wickens, 2016). In total, subjects participated in 27 compensation-relevant tasks. Specifically, the compensation-relevant tasks included the eight private-feedback contests, the eight leaderboard-feedback contests, the risk-aversion elicitation task, the loss-aversion elicitation task, the sunk-cost-elicitation task, and the eight individual search tasks. At the end of the experiment, two of these 27 tasks were chosen at random by the computer for payment.

We recruited 96 students on the campus of Purdue University using ORSEE software (Greiner, 2015). Participants were split into 12 sessions, with eight participants per session. As mentioned above, in order to estimate an average treatment effect over the pooled data, half of the sessions started out with eight private-feedback contests, whereas the other half of the sessions started out with eight leaderboard-feedback contests. The experiment lasted under 60 minutes, with average earnings of \$19.91.

[15] `https://ipip.ori.org/`

4 Predictions

In this section, we present predictions for the experiment that were obtained by solving for the closed-form perfect public equilibrium described in section 2 for the particular model parameters specified in the experiment. In particular, using the model, the resulting predictions were organized into four hypotheses: the first hypothesis pertains to the comparison of the private- and leaderboard-feedback contests; the second hypothesis pertains to the comparison of leader and follower behavior; the third hypothesis pertains to the dynamics of the draws in the two contests; and the fourth hypothesis pertains to the role of individual characteristics such as risk aversion, loss aversion, and the sunk-cost fallacy. Note that these hypotheses are for the particular model parameters specified in the experiment ($v = \$10$, $c = \$1$, $T = 10$, and an exponential distribution of innovation quality with $\lambda = 0.125$), which we chose because they provide interesting qualitative model predictions in a simple environment. Furthermore, it is straightforward to provide examples of parameter configurations that generate qualitatively different model predictions.[16]

[16] For example, if the number of periods, which is set at $T = 10$ in the experiment, becomes arbitrarily large, then the prediction of which level of feedback leads to more draws switches from private-feedback to leaderboard-feedback.

Table 2: **Summary of Predictions**

	Private Feedback	Leaderboard Feedback
Aggregate Draws	8.36	6.34
Proportion of Draws		
Leader		
Known Score 0-15	0.85/0.33/0.08	0.59/0.04/0.00
Known Score > 15	0.00/0.00/0.00	0.00/0.00/0.00
Follower		
Known Score 0-15	0.97/0.66/0.34	0.59/0.55/1.00
Known Score > 15	0.00/0.00/0.00	0.08/0.26/0.21

Notes: Aggregate draws refers to the predicted number of draws that occurs in a contest in each treatment.

Known score refers to the individual score in the private-feedback treatment and the maximum score in the leaderboard-feedback treatment. The third row displays the draw rate of the leader and the follower in periods 2, 6, and 10 of the experiment. The fourth row displays the draw rate in periods 2, 6, and 10 of the experiment for known scores in the 20th-80th percentiles for that period. The fifth row displays the difference in draw rates for known scores in the lower half and the upper half of the known score distribution for periods 2, 6, and 10.

The top part of Table 2 shows that a contest with private feedback is predicted to induce more draws (8.36) than a contest with leaderboard feedback (6.34 draws).[17] We summarize this prediction with Hypothesis 1.[18]

Hypothesis 1 *The private-feedback contest leads to more draws than the leaderboard-feedback*

[17]We used a numerical integration approach to calculate the equilibrium predictions presented in Table 2. In particular, we simulate one million contests for two players following equilibrium strategies derived in Online Appendix A.

[18]In addition, private feedback is predicted to result in a higher winning innovation score (23.42) than a contest with leaderboard feedback (21.84). However, because in the experiment the winning draw will be a noisy estimate of the expected value of the winning innovation, our hypotheses will focus on the decisions to draw rather than the noisy outcome of the draw.

contest.

The bottom part of Table 2 presents the proportion of time players chose to draw an innovation. The proportions are broken down by the period of the contest (presented as a triple of the 2nd/6th/10th periods), the current score (≤ 15 or > 15), and whether the player was a leader or a follower.[19] By comparing the proportion of draws between leaders and followers, the follower is clearly predicted to be at least as likely to draw as the leader across most of the ranges of innovation scores and periods.[20] We summarize this prediction with Hypothesis 2.

Hypothesis 2 *Followers draw more frequently than leaders.*

The bottom part of Table 2 also provides an insight regarding the dynamics of decision-making. In the private-feedback treatment, as the individual innovation score increases, each player becomes less willing to draw. This decrease in willingness to draw can be seen by comparing the proportion of draws between relatively low individual scores (≤ 15) and relatively high individual scores (> 15) for both leaders and followers. Additionally, in the leaderboard-feedback treatment, as the maximum score increases, each player becomes less willing to draw. This can be seen by comparing the proportion of draws between relatively low maximum scores (≤ 15) and relatively high maximum scores (> 15) for both leaders and followers. We summarize this prediction with Hypothesis 3.

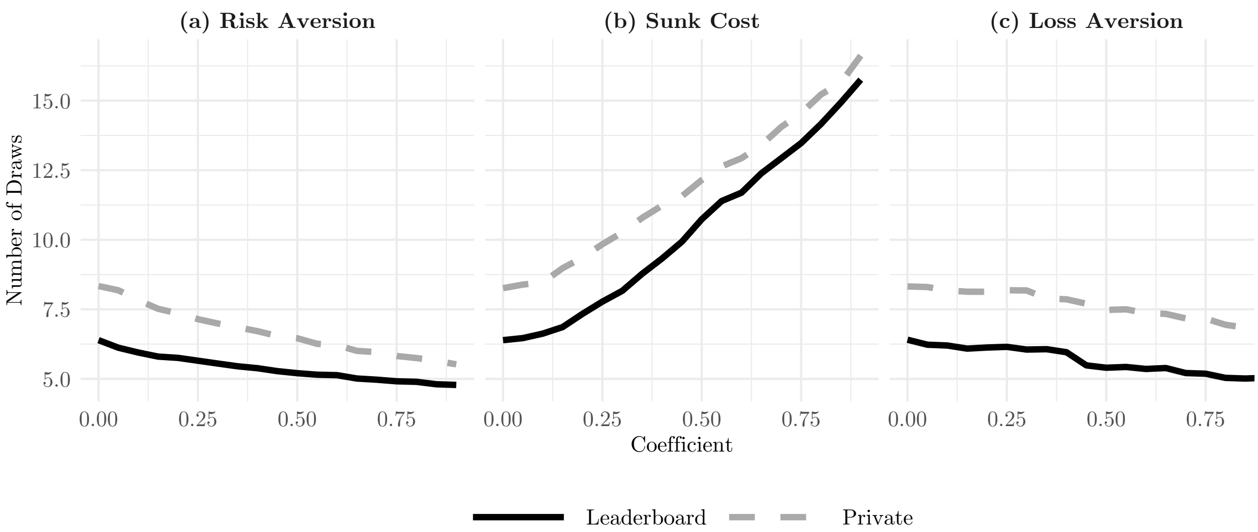
Hypothesis 3 *Players become less willing to draw as their individual score increases in the private-feedback treatment and as the maximum score increases in the leaderboard-feedback treatment.*

[19]Figure D6 in Online Appendix present further evidence on the proportion of draws obtained via numerical integration.

[20]Overall, leaders draw 8.73% of the time in the simulated contests and followers draw 39.20% of the time in the simulated contests.

Lastly, we incorporate three behavioral characteristics: risk aversion, loss aversion, and the sunk-cost fallacy.[21] The three panels of Figure 4 present the comparative statics as we vary these characteristics one at a time. For example, to vary risk aversion, we model both players as having a CRRA utility function with parameter γ, and we vary this parameter across a range of values typically observed in the experimental literature.

Figure 4: **Decision to Draw and Comparative Statics**



Notes: This figure displays equilibrium predictions under different levels of (a) risk aversion, (b) the sunk-cost fallacy, and (c) loss aversion. The dashed gray line is the private-feedback treatment, and the solid black line is the leaderboard-feedback treatment.

Figure 4 shows that as risk aversion and loss aversion increase, the number of total draws made in the contest decreases. The sunk-cost fallacy, however, has an opposite effect. In particular, as the sunk-cost fallacy increases, we observe more total draws. We summarize these predictions with Hypothesis 4.

[21]Specifications of the three utility functions as well as the general procedure for obtaining predictions are provided in Online Appendix B.

Hypothesis 4 *The number of draws increases with (a) a decrease in risk aversion, (b) a decrease in loss aversion, and (c) an increase in the sunk-cost fallacy.*

5 Results

In this section, we present the results of our experiment. In particular, first, in section 5.1 we compare the outcomes of the private and leaderboard treatments. Next, in section 5.2, we test for differences in behavior between the leader and the follower. Then, in section 5.3, we consider the dynamics observed in the experimental data. Finally, in section 5.4, we discuss the role of individual characteristics in determining innovation-contest outcomes.

5.1 Private vs Leaderboard Contests

Table 3 presents the summary statistics from the two treatments. In particular, the table is divided into two parts. In the top part, we present the aggregate results on the total number of draws that we observed in each of the treatments, on average. In the bottom part, we present the results on the proportion of draws conditional on the period in the game (periods 2, 6, and 10 are separated by "/"), current score, and whether the decision-maker was a leader or a follower.[22]

[22]Recall that although the role of leader/follower is known to the decision-makers in the leaderboard-feedback treatment, it is not known to the decision-makers in the private-feedback treatment.

Table 3: **Contest Results**

	Private Feedback	Leaderboard Feedback
Aggregate Draws	8.50	7.54
Proportion of Draws		
Leader		
Known Score 0-15	0.81/0.42/0.33	0.61/0.21/0.20
Known Score > 15	0.28/0.13/0.09	0.15/0.04/0.06
Follower		
Known Score 0-15	0.79/0.45/0.42	0.70/0.48/0.63
Known Score > 15	0.40/0.12/0.21	0.56/0.37/0.39

Notes: Aggregate draws refers to the observed mean number of draws that occur in a contest in each treatment.

The third row displays the draw rate of the leader and the follower in periods 2, 6, and 10 of the experiment. The fourth row displays the draw rate in periods 2, 6, and 10 of the experiment for scores that range in the 20th-80th percentiles for that period. The fifth row displays the difference in draw rates for scores in the lower half and the upper half of the score distribution for periods 2, 6, and 10.

The top part of Table 3 shows the average number of contest draws in each treatment.In particular, in the private-feedback treatment, the average number of draws (8.50) is not significantly different from the theoretically predicted value (8.36 draws, p-value 0.67).[23] In terms of the leaderboard feedback, we do find a difference between theory and the experiment in terms of the number of draws for the leaderboard-feedback treatment (6.34 vs. 7.54, *p*-value <0.01).[24]

[23]Hypothesis tests in this paragraph are conducted using bootstrapped regressions, with 5,000 bootstrap samples, on the session-level averages.

[24]When testing hypotheses, we use individual-level data when the dependent variable is determined solely by an individual (e.g., individual decision to make an additional innovation draw). When the dependent variable is determined by multiple individuals, we use a higher level. For example, when testing the treatment difference between aggregate draws, we use contest-level data with each 10-period contest as a unit of analysis.

The main focus of the aggregate results is on the comparison between private and leaderboard feedback (i.e., Hypothesis 1). Table 3 shows that in our experiment, the number of draws in the private-feedback contest (8.50) is greater than in the leaderboard-feedback contest (7.54).[25] We test whether this difference is significant using a fixed-effects regression with session-level effects.[26] We find that this difference is significant (p-value<0.01).[27] Table D4 in the Online Appendix shows that this conclusion is robust when we control for the order in which the two contests were presented as well as when we restrict the analysis to the first contest faced by the participant. We summarize these tests with Result 1.

Result 1 *A private-feedback contest results in more draws than a leaderboard-feedback contest (evidence supporting Hypothesis 1).*

5.2 Leaders vs. Followers

The bottom part of Table 3 shows that the proportion of time that a follower draws is greater than the proportion of time that a leader draws. Although the difference is observed in both the private and leaderboard treatments, the difference is much larger in the latter. Figure 5 presents further evidence regarding this comparison. Formally, each panel of the figure shows a panel data logistic regression of the decision to draw on the maximum score. The bottom row of the figure presents the comparison of the leader's decision (solid black) and the follower's decision (dashed gray). The figure clearly shows that in almost every combination of period and maximum score, followers are more likely to draw than leaders.

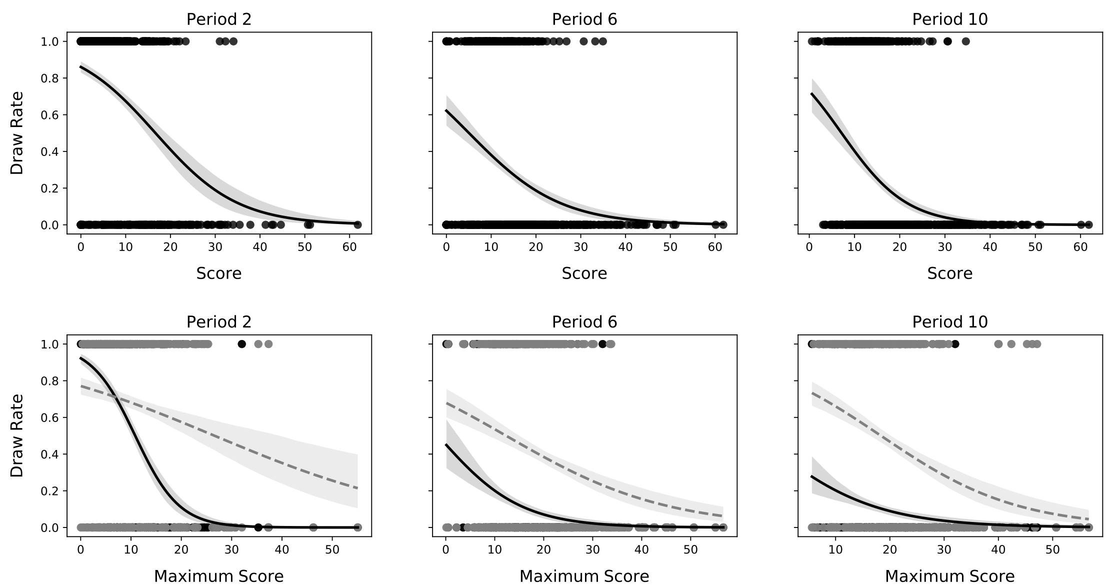
[25]The private-feedback contest also results in a higher winning innovation, on average. The private-feedback contest results in a significantly higher winning innovation at the five percent level (p-value=0.03) using a fixed effects regression with session-level effects.

[26]We can also test whether this difference is significant using non-parametric tests. A Wilcoxon Signed-rank test using the session-level (subject-level) average number of draws in each treatment is significant at the five (one) percent level.

[27]The p-value<0.01 if we utilize a random-effects regression with session-level effects.

Thus, Figure 5 suggests that Hypothesis 2 holds.[28]

Figure 5: **Decision to Draw in the Leaderboard-Feedback Treatment**



Notes: This figure displays two sets of graphs. The first row display logistic regressions of the decision to draw in the private-feedback treatment for periods 2, 6, and 10. The second row display logistic regressions of the leader's decision (solid black) to draw and the follower's decision (dashed gray) to draw in the leaderboard-feedback treatment for periods 2, 6, and 10.

To formally test the difference between leader and follower behavior, we use a panel data logistic regression. In particular, we regress the decision to draw on an indicator variable for whether the subject was a leader, while accounting for subject-level fixed effects and clustering standard errors at the session level.[29] The coefficient on the leader variable is negative and significant at the 1% level. We summarize these observations with Result 2.

[28]Figure D5 in the Online Appendix provides similar figures for the remaining periods.

[29]Note that the regression is run on the observations where the score is greater than zero (and thus there is a leader and a follower).

Result 2 *Leaders draw less frequently than followers in the leaderboard-feedback treatment (evidence supporting Hypothesis 2).*

5.3 Dynamics of Decision-Making

Figure 5 suggests that subjects are less willing to draw as the individual score increases in the private-feedback treatment and as the maximum score increases in the leaderboard-feedback treatment. To formally test Hypothesis 3, we run panel data logistic regressions, with subject-level fixed effects and session-level clustered standard errors, of the decision to draw on the individual score. We run these regressions for the last nine periods of the private-feedback treatment. We find that in each of the regressions, the coefficient on the individual score is negative and significant at the 1% level. Additionally, we run similar regressions for the leaderboard-feedback treatment, with the difference being that the decision to draw is regressed on the maximum score. Again, for each of the regressions, the coefficient on the maximum score is negative and significant at the 1% level. We summarize these results with Result 3.

Result 3 *Subjects are less willing to draw as their individual score increases in the private-feedback treatment and as the maximum score increases in the leaderboard-feedback treatment (evidence supporting Hypothesis 3).*

5.4 Role of Individual Characteristics

In our experiment, subjects completed various elicitation tasks. We used these tasks to shed light on factors that may influence subjects' decision to draw. Table 4 displays three sets of regressions that analyze the decision to draw on the elicited characteristics.[30] In particular, the regressions are carried out using a panel data logistic regression with subject-level random effects, and standard errors are obtained by clustering at the session level.

Table 4 shows that the regression analyses yield results consistent with our prior analysis in terms of the role of the treatments and leader/follower behavior. In terms of elicited

[30]Results for the individual search task are similar (see regression results presented in Table D2 of the Online Appendix).

individual characteristics, we find that risk aversion has a significantly negative effect across a number of specifications.[31] At the same time, we find that our measures of loss aversion and sunk-cost fallacy are not significant in any of the specifications. We summarize these results with Result 4.

Result 4 *Risk aversion leads to a lower likelihood of drawing an innovation (evidence supporting Hypothesis 4a).*

Recall that in addition to the incentivized elicitation of risk aversion, loss aversion, and the sunk-cost fallacy, we conducted a number of non-incentivized personality questionnaires that addressed personality characteristics. In particular, in addition to a broad questionnaire (i.e., Big 5), we selected a few characteristics as potentially important to behavior in an innovation-contest setting (i.e., Grit and Competitiveness). Table 4 shows that virtually no personality characteristics are significant in explaining drawing behavior for any of the regression specifications.

[31]Although, at the aggregate level, risk aversion was predicted and was observed to have negative impact on the proportion of draws, heterogeneity in risk aversion (i.e., risk loving subjects) in combination with heterogeneity in other behavioral (e.g., loss aversion) and psychological characteristics (e.g., competitiveness) as well as noisy decision making may lead to appearance of excessive draws when comparing Table 2 and Table 3 (especially in the last period in which the prediction is at the boundary).

Table 4: **Regression Results**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dep. Var.:	Pooled		Private			Leaderboard	
Draw Decision		All	Leader	Follower	All	Leader	Follower
L-Board	-0.25***	—	—	—	—	—	—
	(0.08)	—	—	—	—	—	—
Individual Score	—	-0.21***	-0.25***	-0.18***	—	—	—
	—	(0.02)	(0.04)	(0.02)	—	—	—
Maximum Score	—	—	—	—	-0.11***	-0.23***	-0.11***
	—	—	—	—	(0.01)	(0.02)	(0.01)
Period	-0.30***	-0.13***	-0.19***	-0.11***	-0.10***	-0.24***	-0.03
	(0.02)	(0.03)	(0.04)	(0.04)	(0.03)	(0.04)	(0.05)
Risk Aversion	-0.70**	-1.41**	-1.50	-1.20**	-1.05**	-1.01	-0.31
	(0.31)	(0.72)	(1.32)	(0.56)	(0.46)	(0.87)	(1.15)
Loss Aversion	0.02	-0.10	1.15	-0.83	-0.30	-1.12	-0.43
	(0.51)	(0.83)	(1.02)	(0.70)	(0.63)	(1.09)	(0.89)
Sunk Cost Fallacy	-0.07	0.14	-1.07	0.25	-0.12	-0.55	0.02
	(0.38)	(0.94)	(0.87)	(0.96)	(0.45)	(0.87)	(0.94)
Grit	-0.04	-0.15	-0.29	-0.03	-0.01	-0.09	-0.11
	(0.11)	(0.21)	(0.25)	(0.18)	(0.08)	(0.22)	(0.25)
Competitiveness	-0.07	0.07	0.00	0.24	-0.24	-0.04	-0.15
	(0.11)	(0.24)	(0.26)	(0.23)	(0.16)	(0.21)	(0.20)
Achievement Striving	0.08	0.08	0.03	-0.09	0.27	0.31	0.04
	(0.13)	(0.26)	(0.32)	(0.24)	(0.17)	(0.22)	(0.33)
Extraversion	0.07	-0.03	0.07	-0.06	0.08	-0.18	0.11
	(0.07)	(0.11)	(0.14)	(0.10)	(0.10)	(0.19)	(0.11)
Agreeableness	0.09	0.06	0.02	0.00	0.16	0.12	0.16
	(0.10)	(0.17)	(0.22)	(0.18)	(0.14)	(0.20)	(0.20)
Neuroticism	0.02	0.05	-0.12	0.10	0.03	0.11	-0.07
	(0.08)	(0.14)	(0.18)	(0.12)	(0.11)	(0.18)	(0.22)
Openness	-0.07	-0.11	-0.16	-0.14	-0.13	-0.27	-0.14
	(0.07)	(0.15)	(0.19)	(0.14)	(0.09)	(0.19)	(0.15)
Conscientiousness	-0.03	0.18	0.25	0.06	-0.14	-0.20	-0.24
	(0.11)	(0.29)	(0.27)	(0.26)	(0.10)	(0.35)	(0.20)
Constant	1.03**	1.98**	4.41***	1.32	1.00*	2.01**	1.75**
	(0.42)	(0.81)	(0.82)	(0.85)	(0.51)	(0.83)	(0.69)
Observations	15,360	7,680	3,451	3,451	7,680	3,411	3,411

Notes: The regression pools the data from the private-feedback treatment and the leaderboard-feedback treatment. Personality characteristics are standardized to have mean 0.00 and standard deviation of 1.00. *,** , and *** denote significance at the 0.10, 0.05, and 0.01 levels, respectively.

6 Additional Treatments

An important question is whether our theoretical and experimental results hold if we increase the number of players. After all, sequential-sampling competitions typically involve more than two players. To address this concern, we extended the theory to handle an arbitrary number of players (N) and conducted a set of new experiments with four players

($N = 4$). As in the two-player case, the theoretical analysis focuses on a game between two types of players – leaders and followers. As before, each stage game at $t > 1$ has only one leader. However, unlike the case with two players, there are now more than one follower. What makes the theory tractable is the fact that all followers are symmetric in that they have the same information and are drawing from the same distribution.[32] We use the extended models to compute the perfect public equilibrium for four-player sequential-sampling contests of two types – those with public-leaderboard feedback and those with private feedback – and provide predictions for the new experiments described next.

Regarding the new experiments, our goal was to make them as close as possible to the original experiments discussed in section 3. In particular, we kept the same within-subject experimental design whereby each subject faced 27 compensation-relevant tasks. In terms of the parameters of the contests, we kept the same distribution of innovation quality ($F(x; \lambda) = 1 - e^{-\lambda x}$ with $\lambda = 0.125$), the same same finite time horizon ($T = 10$), the same initial endowment of 10.00, and the same cost of drawing ($c = \$1.00$). The two changes that we made included increasing the number of players within each contest from two to four and increasing the contest prize from $v = \$10.00$ to $v = \$30.00$. The increase in the prize was motivated by the fact that with more players, each one is less likely to win the prize and instead is more likely to face low earnings within a contest due to the cost of drawing.

[32]For details on the model with $N > 2$ players, see Online Appendix A.

Table 5: **Summary of Predictions (Additional Treatments, $N = 4$)**

	Private Feedback	Leaderboard Feedback
Aggregate Draws	26.17	21.40
Proportion of Draws		
Leader		
Known Score 0-15	1.00/1.00/1.00	1.00/1.00/0.56
Known Score > 15	0.30/0.09/0.02	0.39/0.10/0.00
Follower		
Known Score 0-15	1.00/1.00/1.00	1.00/1.00/1.00
Known Score > 15	0.57/0.42/0.31	0.74/0.53/0.34

Notes: Aggregate draws refers to the predicted number of draws that occurs in a contest in each treatment. Known score refers to the individual score in the private-feedback treatment and the maximum score in the leaderboard-feedback treatment. The third row displays the draw rate of the leader and the follower in periods 2, 6, and 10 of the experiment. The fourth row displays the draw rate in periods 2, 6, and 10 of the experiment for known scores in the 20th-80th percentiles for that period. The fifth row displays the difference in draw rates for known scores in the lower half and the upper half of the known score distribution for periods 2, 6, and 10.

Table 5 presents theoretical predictions for the new set of parameters. The table shows that (i) the private feedback is predicted to induce more draws (26.17) than the leaderboard feedback (21.40); (ii) followers draw more frequently than leaders (e.g., the proportion of time that a follower draws in period 10 when the known score is greater than 15 is 0.34 vs. 0.00 for the leader); and (iii) subjects are less willing to draw as their individual score increases in the private-feedback treatment and the maximum score increases in the leaderboard-feedback treatment (e.g., the proportion of time that a follower draws in period 10 when the known score is greater than 15 is 0.34 vs. 1.00 when the known score is between 0 and 15). These predictions directly match Hypotheses 1–3 discussed in section 4.

To test these predictions, we ran a new experiment with 172 students recruited on the campus of Purdue University who had not participated in the original experiments. As before, to ensure that the order of private- and leaderboard-feedback treatments did not affect the results, half of the sessions started with eight private-feedback contests, and the other half started with eight leaderboard-feedback contests. Table 6 presents the summary of results from our experiments.

Table 6: **Contest Results (Additional Treatments, $N = 4$)**

	Private Feedback	Leaderboard Feedback
Aggregate Draws	18.83	14.93
Proportion of Draws		
Leader		
Known Score 0-15	0.78/0.43/0.33	0.64/0.42/0.13
Known Score > 15	0.34/0.23/0.18	0.17/0.12/0.12
Follower		
Known Score 0-15	0.72/0.46/0.41	0.71/0.51/0.60
Known Score > 15	0.41/0.30/0.30	0.43/0.27/0.29

Notes: Aggregate draws refers to the number of draws made in a contest in each treatment. The third row displays the draw rate of the leader and the follower in periods 2, 6, and 10 of the experiment. The fourth row displays the draw rate in periods 2, 6, and 10 of the experiment for scores that range in the 20th-80th percentiles for that period. The fifth row displays the difference in draw rates for scores in the lower half and the upper half of the score distribution for periods 2, 6, and 10.

We find strong support for all of the hypotheses discussed in section 4. Specifically, we find that the number of aggregate draws is significantly higher in the private-feedback treatment than in the leaderboard-feedback treatment (18.83 vs. 14.93; p-value $< .01$ using a fixed-effects regression with session-level fixed effects) providing support for Hypothesis 1. Regarding Hypothesis 2, we find that followers are significantly more likely to draw than the the leaders (p-value $< .01$ using fixed-effects logistic regression of the decision to draw on an indicator variable for whether the subject was a leader). Finally, we find strong support for Hypothesis 3, namely, that subjects are less willing to draw as their individual (maximum) score increases in the private-feedback (leaderboard-feedback) treatment (p-value $< .01$ using panel data logistic regressions, with subject-level fixed effects and session-level clustered standard errors).

In addition to the main results, we also find partial support for Hypothesis 4 that considers the individual characteristics (i.e., risk aversion, loss aversion, and sunk-cost fallacy). Specifically, as in the main experiment, we find strong support that individual risk aversion leads to a lower likelihood of drawing an innovation (p-value $< .01$ using random-effects regressions with subject-level random effects and session-level clustered standard errors). We also find support that the sunk-cost fallacy increases the likelihood to draw. As one would expect, the increase is driven by the followers and not the leaders. Interestingly, this result was not present in the two-player experiments, indicating the sunk-cost fallacy is relevant with more players. As with two-player experiments, we did not find loss aversion to be a predictor of behavior in our contests. Lastly, among the personality characteristics, in the new sets of experiments, we find that an unincentivized measure of competitiveness was correlated with a higher propensity of the followers to draw in an attempt to overtake the leader.[33]

[33] In the Online Appendix we consider alternative specifications. In particular, in Table D7 we control for both the individual score (privately observed in both treatments) and maximum score (observed in the leaderboard-feedback treatment but not in the private-feedback treatment), while in Table D8 (and Figure D7) we interact treatment dummy with individual-level covariates. Our main results are robust to these alternative specifications.

Table 7: **Regression Results**

Dep. Var.: Draw Decision	(1) Pooled	(2) Private All	(3) Private Leader	(4) Private Follower	(5) Leaderboard All	(6) Leaderboard Leader	(7) Leaderboard Follower
L-Board	-0.54***	—	—	—	—	—	—
	(0.10)	—	—	—	—	—	—
Individual Score	—	-0.13***	-0.18***	-0.10***	—	—	—
	—	(0.01)	(0.02)	(0.01)	—	—	—
Maximum Score	—	—	—	—	-0.11***	-0.21***	-0.12***
	—	—	—	—	(0.01)	(0.04)	(0.01)
Period	-0.29***	-0.14***	-0.15***	-0.16***	-0.08***	-0.14***	-0.08***
	(0.02)	(0.02)	(0.03)	(0.03)	(0.02)	(0.04)	(0.02)
Risk Aversion	-2.84***	-4.07***	-3.64*	-3.77***	-3.64***	-3.38	-4.38***
	(0.60)	(1.06)	(1.92)	(1.14)	(0.83)	(2.42)	(1.01)
Loss Aversion	0.06	-0.01	0.64	0.34	-0.07	0.98	-0.29
	(0.39)	(0.50)	(0.66)	(0.73)	(0.73)	(0.93)	(1.00)
Sunk Cost Fallacy	1.27***	2.19***	-0.07	2.42***	1.39**	0.75	1.77**
	(0.43)	(0.69)	(0.64)	(0.81)	(0.61)	(1.15)	(0.90)
Grit	-0.03	-0.17	0.05	-0.20	0.07	0.03	0.04
	(0.11)	(0.20)	(0.24)	(0.21)	(0.12)	(0.22)	(0.20)
Competitiveness	0.20**	0.28**	-0.06	0.36**	0.31***	-0.12	0.37***
	(0.09)	(0.14)	(0.18)	(0.15)	(0.10)	(0.18)	(0.11)
Achievement Striving	-0.13	-0.20	0.08	-0.25*	-0.17*	0.19	-0.20*
	(0.08)	(0.14)	(0.18)	(0.14)	(0.09)	(0.28)	(0.11)
Extraversion	-0.01	-0.00	0.39***	-0.09	-0.07	-0.13	-0.04
	(0.10)	(0.12)	(0.14)	(0.15)	(0.13)	(0.20)	(0.16)
Agreeableness	0.13	0.16	-0.26	0.11	0.17	0.02	0.15
	(0.12)	(0.20)	(0.22)	(0.21)	(0.13)	(0.16)	(0.15)
Neuroticism	0.05	0.13	0.05	0.13	0.04	-0.12	0.03
	(0.07)	(0.11)	(0.12)	(0.13)	(0.10)	(0.24)	(0.12)
Openness	0.02	0.14	0.26	0.18	-0.11	0.08	-0.09
	(0.10)	(0.14)	(0.17)	(0.13)	(0.14)	(0.17)	(0.16)
Conscientiousness	-0.04	-0.13	0.01	-0.14	0.07	-0.24	0.06
	(0.21)	(0.37)	(0.37)	(0.40)	(0.21)	(0.30)	(0.34)
Constant	-0.69*	-1.08*	2.01	-1.03*	-0.78	0.94	-0.78
	(0.36)	(0.57)	(1.44)	(0.62)	(0.50)	(1.52)	(0.61)
Observations	27,520	13,760	3,094	9,282	13,760	3,088	9,264

Notes: The regression pools the data from the the private-feedback treatment and the leaderboard-feedback treatment. Personality characteristics are standardized to have mean 0.00 and standard deviation of 1.00. *,** , and *** denote significance at the 0.10, 0.05, and 0.01 levels, respectively.

To summarize, we ran additional experiments with $N = 4$ players and find results that replicate the original findings remarkably well. In particular, our results hold both in terms of the aggregate number of draws (greater number of draws in the private-feedback treatment than in the leaderboard treatment) and the drawing dynamics (followers drawing more than leaders, and the propensity to draw decreasing in the best-known value of innovation). In addition, we find that the individual incentivized measure of risk aversion is a strong predictor of the decision to draw. Among the differences observed between the two sets of experiments, we find that when the number of players increases, the role of the sunk-cost fallacy and competitiveness becomes more important.

7 Conclusion

In this paper, we investigate the role of leaderboard feedback in sequential-search innovation competition. In particular, our contribution is threefold. First, we contribute to the experimental literature that investigates dynamic contests and innovation competitions. Our experiment yields several results that support the theoretical predictions. Specifically, we find that for a two-player finite-horizon contest, leaderboard feedback may yield less effort and lower innovation quality than private feedback. We also find that the internal dynamics present in the data are consistent with the model. In particular, when feedback is provided, leaders of the contest reduce their effort, whereas followers do not. In addition, as the quality of innovation increases, agents become less likely to invest resources to generate a new innovation.

Second, our work also contributes to a stream of literature that studies the role of individual characteristics in determining an individual's propensity to innovate. In particular, we elicit three individual characteristics that have been shown to be important in the innovation and contest setting: risk aversion, loss aversion, and the sunk-cost fallacy. We find that among these individual characteristics, risk aversion is a consistent driver of behavior in both the two-player and four-player contests. At the same time, loss aversion is not significant in explaining the data in either setting. The sunk-cost fallacy turns out to matter only when the number of players is higher and only for followers. This makes sense, because with a greater number of players, each one is more likely to draw without ever becoming a leader, creating conditions that are inducing sunk-cost fallacy considerations. In addition, we find little evidence that personality characteristics are predictive of behavior in the dynamic contests studied in this paper.

Finally, we contribute to the existing theoretical literature by examining equilibrium in a model of sequential-sampling competition with a finite horizon and imperfect public monitoring. We find that with a finite horizon, leaderboard feedback may result in lower search effort as captured by the number of costly innovation decisions, which in turn yields a lower expected quality of the winning innovation with leaderboard feedback than with private feedback.

Our work has several shortcomings that open interesting avenues for future research. First, our paper investigates a finite-horizon innovation competition. Comparing it to with an infinite-horizon setting would be interesting. Second, our experiment had individuals make innovation decisions on their own. It would be interesting to see how innovation decisions differ when groups are making these decisions. Third, subjects in our experiment participated in the contest (although they had an option not to draw). Investigating the extent to which our results hold if subjects could select to withdraw from the contests entirely would be interesting. Finally, in this paper, we considered winner-take-all contests. The extent to which the main results translate to shared-prize contests is not known.

References

Ales, Laurence, Soo-Haeng Cho, and Ersin Körpeoğlu. 2017. "Optimal award scheme in innovation tournaments." *Operations Research*, 65(3): 693–702.

Aoyagi, Masaki. 2010. "Information feedback in a dynamic tournament." *Games and Economic Behavior*, 70(2): 242–260.

Astebro, Thomas, Holger Herz, Ramana Nanda, and Roberto A Weber. 2014. "Seeking the roots of entrepreneurship: Insights from behavioral economics." *Journal of Economic Perspectives*, 28(3): 49–70.

Augenblick, Ned. 2015. "The sunk-cost fallacy in penny auctions." *The Review of Economic Studies*, 83(1): 58–86.

Banovetz, James, and Ryan Oprea. 2022. "Complexity and Procedural Choice." *Work-*

ing Paper.

Baye, Michael R, and Heidrun C Hoppe. 2003. “The strategic equivalence of rent-seeking, innovation, and patent-race games.” *Games and economic behavior*, 44(2): 217–226.

Berry, Donald A. 1972. “A Bernoulli two-armed bandit.” *Annals of Mathematical Statistics*, 43: 871–897.

Bimpikis, Kostas, Shayan Ehsani, and Mohamed Mostagir. 2019. “Designing dynamic contests.” *Operations Research*, 67(2): 339–356.

Bonatti, Alessandro, and Johannes Hörner. 2011. “Collaborating.” *American Economic Review*, 101: 632–663.

Bradt, Russell N, and Samuel Karlin. 1956. “On the design and comparison of certain dichotomous experiments.” *Annals of Mathematical Statistics*, 27: 390–409.

Chawla, Shuchi, Jason D Hartline, and Balasubramanian Sivan. 2015. “Optimal crowdsourcing contests.” *Games and Economic Behavior.*

Chen, Daniel L, Martin Schonger, and Chris Wickens. 2016. “oTree—An open-source platform for laboratory, online, and field experiments.” *Journal of Behavioral and Experimental Finance*, 9: 88–97.

Che, Yeon-Koo, and Ian Gale. 2003. “Optimal design of research contests.” *American Economic Review*, 93(3): 646–671.

Dechenaux, Emmanuel, Dan Kovenock, and Roman M Sheremeta. 2015. “A survey of experimental research on contests, all-pay auctions and tournaments.” *Experimental Economics*, 18(4): 609–669.

Deck, Cary, and Erik O Kimbrough. 2017. “Experimenting with contests for experimentation.” *Southern Economic Journal*, 84(2): 391–406.

DeGroot, Morris H. 1970. *Optimal Statistical Decisions.* Wiley New York.

DiPalantino, Dominic, and Milan Vojnovic. 2009. "Crowdsourcing and all-pay auctions." 119–128, ACM.

Duckworth, Angela L, Christopher Peterson, Michael D Matthews, and Dennis R Kelly. 2007. "Grit: perseverance and passion for long-term goals." *Journal of personality and social psychology*, 92(6): 1087.

Ederer, Florian, and Gustavo Manso. 2013. "Is pay for performance detrimental to innovation?" *Management Science*, 59(7): 1496–1513.

Erat, Sanjiv, and Vish Krishnan. 2012. "Managing delegated search over design spaces." *Management Science*, 58(3): 606–623.

Feldman, Dorian. 1962. "Contributions to the "two-armed bandit" problem." *Annals of Mathematical Statistics*, 33: 847–856.

Fershtman, Chaim, and Uri Gneezy. 2011. "The tradeoff between performance and quitting in high power tournaments." *Journal of the European Economic Association*, 9(2): 318–336.

Fudenberg, Drew, Phillip Strack, and Tomasz Strzalecki. 2018. "Speed, Accuracy, and the Optimal Timing of Choices." *American Economic Review*, 108: 3651–3684.

Fullerton, Richard L, and R Preston McAfee. 1999. "Auctionin entry into tournaments." *Journal of Political Economy*, 107(3): 573–605.

Gershkov, Alex, and Motty Perry. 2009. "Tournaments with midterm reviews." *Games and Economic Behavior*, 66(1): 162–190.

Goltsman, Maria, and Arijit Mukherjee. 2011. "Interim performance feedback in multistage tournaments: The optimality of partial disclosure." *Journal of Labor Economics*, 29(2): 229–265.

Greiner, Ben. 2015. "Subject pool recruitment procedures: organizing experiments with ORSEE." *Journal of the Economic Science Association*, 1(1): 114–125.

Gross, Daniel P. 2017. "Performance feedback in competitive product development."

RAND Journal of Economics, 48: 438–466.

Halac, Marina, Navin Kartik, and Qingmin Liu. 2016. "Optimal Contracts for Experimentation." *Review of Economic Studies*, 83: 1040–1091.

Halac, Marina, Navin Kartik, and Qingmin Liu. 2017. "Contests for experimentation." *Journal of Political Economy*, 125(5): 1523–1569.

Herz, Holger, Daniel Schunk, and Christian Zehnder. 2014. "How do judgmental overconfidence and overoptimism shape innovative activity?" *Games and Economic Behavior*, 83: 1–23.

Hinnosaar, Toomas. 2016. "Penny auctions." *International Journal of Industrial Organization*, 48: 59–87.

Hoelzemann, Johannes, and Nicolas Klein. 2021. "Bandits in the lab." *Quantitative Economics*, 12: 1021–1051.

Holt, Charles A, and Susan K Laury. 2002. "Risk aversion and incentive effects." *American economic review*, 92(5): 1644–1655.

Hudja, Stanton. 2021. "Is experimentation invariant to group size? A laboratory analysis of innovation contests." *Journal of Behavioral and Experimental Economics*, 91: 101660.

John, Oliver P, and Sanjay Srivastava. 1999. "The Big Five trait taxonomy: History, measurement, and theoretical perspectives." *Handbook of personality: Theory and research*, 2(1999): 102–138.

Keller, Godfrey, Sven Rady, and Martin Cripps. 2005. "Strategic Experimentation with Exponential Bandits." *Econometrica*, 73: 39–68.

Khromenkova, Daria. 2015. "Collective Experimentation with Breakdowns and Breakthroughs." *Working Paper.*

Klein, Arnd Heinrich, and Armin Schmutzler. 2017. "Optimal effort incentives in dynamic tournaments." *Games and Economic Behavior*, 103: 199–224.

Koudstaal, Martin, Randolph Sloof, and Mirjam Van Praag. 2015. "Risk, uncertainty, and entrepreneurship: Evidence from a lab-in-the-field experiment." *Management Science*, 62(10): 2897–2915.

Kuhnen, Camelia M, and Agnieszka Tymula. 2012. "Feedback, self-esteem, and performance in organizations." *Management Science*, 58(1): 94–113.

Lang, Matthias, Christian Seel, and Philipp Strack. 2014. "Deadlines in stochastic contests." *Journal of Mathematical Economics*, 52: 134–142.

Lemus, Jorge, and Guillermo Marshall. 2021. "Dynamic tournament design: evidence from prediction contests." *Journal of Political Economy*, 129: 383–420.

Ludwig, Sandra, and Gabriele K Lünser. 2012. "Observing your competitor–The role of effort information in two-stage tournaments." *Journal of Economic Psychology*, 33(1): 166–182.

Mailath, George J., and Larry Samuelson. 2006. *Repeated Games and Reputations: Long-Run Relationships.* Oxford University Press Oxford.

Malueg, David A., and Andrew J. Yates. 2010. "Testing contest theory: evidence from best-of-three tennis matches." *The Review of Economics and Statistics*, 92: 689–692.

Mihm, Jürgen, and Jochen Schlapp. 2018. "Sourcing innovation: On feedback in contests." *Management Science*, 65(2): 559–576.

O'Neill, Barry. 1986. "International escalation and the dollar auction." *Journal of Conflict Resolution*, 30(1): 33–50.

Rieck, Thomas. 2010. "Information disclosure in innovation contests." *Bonn Graduate School of Economics Working paper.*

Robbins, Herbert. 1952. "Some aspects of the sequential design of experiments." *Bulletin of the American Mathematical Society*, 58: 527–535.

Rosokha, Yaroslav, and Kenneth Younge. 2020. "Motivating innovation: The effect of loss aversion on the willingness to persist." *Review of Economics and Statistics*, 102(3): 569–582.

Rothschild, Michael. 1974. "A two-armed bandit theory of market pricing." *Journal of Economic Theory*, 9: 185–202.

Rubin, Jared, Anya Samek, and Roman M Sheremeta. 2018. "Loss aversion and the quantity–quality tradeoff." *Experimental Economics*, 21(2): 292–315.

Seel, Christian, and Philipp Strack. 2016. "Continuous time contests with private information." *Mathematics of Operations Research*, 41(3): 1093–1107.

Shubik, Martin. 1971. "The dollar auction game: A paradox in noncooperative behavior and escalation." *Journal of conflict Resolution*, 15(1): 109–111.

Stigler, George J. 1961. "The economics of information." *Journal of Political Economy*, 69: 213–225.

Strulovici, Bruno. 2010. "Learning while voting: Determinants of collective experimentation." *Econometrica*, 78(3): 933–971.

Taylor, Curtis R. 1995. "Digging for golden carrots: An analysis of research tournaments." *The American Economic Review*, 872–890.

Terwiesch, Christian, and Yi Xu. 2008. "Innovation contests, open innovation, and multiagent problem solving." *Management science*, 54(9): 1529–1543.

Wald, Abraham. 1947. *Sequential Analysis.* Wiley New York.

Yildirim, Huseyin. 2005. "Contests with multiple rounds." *Games and Economic Behavior*, 51(1): 213–227.