

# Performance feedback in competitive product development

Daniel P. Gross\*

*Performance feedback is ubiquitous in competitive settings where new products are developed. This article introduces a fundamental tension between incentives and improvement in the provision of feedback. Using a sample of 4294 commercial logo design tournaments, I show that feedback reduces participation but improves the quality of subsequent submissions, with an ambiguous effect on high-quality output. To evaluate this trade-off, I develop a procedure to estimate agents' effort costs and simulate counterfactuals under alternative feedback policies. The results suggest that feedback on net increases the number of high-quality ideas produced and is thus desirable for a principal seeking innovation.*

Performance feedback is pervasive in industries where new products are developed: workers pitch ideas to managers and clients, who provide critiques and direction; prototypes are tested at Research and Development (R&D) labs, with focus groups, and in public demonstrations; and consumers are the ultimate arbiters of value. Despite widespread use, the effects of feedback on innovation in competitive environments are not fully understood: feedback is argued to be essential to improving innovation in single-agent contracts (Manso, 2011), but research on the effects of performance disclosure in routine competitive settings (e.g., Ederer, 2010) has shown that it can weaken incentives by revealing asymmetries between competitors.

This article studies the tension between incentives and improvement in feedback provision, which is of fundamental importance to firms in creative or research industries and other organizations procuring or funding innovation. This tension is intrinsic to product development and is especially stark in formal competitions such as innovation contests (e.g., Taylor, 1995; Che and Gale, 2003; Terwiesch and Xu, 2008), which are undergoing a renaissance in the private,

---

\* Harvard Business School and NBER; [dgross@hbs.edu](mailto:dgross@hbs.edu).

I thank Ben Handel, Przemek Jeziorski, Gustavo Manso, John Morgan, Denis Nekipelov, and the Editor and referees for comments that improved this article, as well as to the Harvard Business School Division of Research and Faculty Development and NSF Graduate Research Fellowship grant no. DGE-1106400 for financial support. Limited short segments of the text may be similar to passages from another of my articles (“Creativity under Fire: The Effects of Competition on Creative Production”), which uses a different data set from the same setting. A previous version of this article circulated under the title “Interim Feedback in Tournaments for Ideas: Trading Off Participation for Quality.” A online Appendix is available at <http://www.dpgross.com/>. All errors are my own.

public, and nonprofit domains (Williams, 2012).<sup>1</sup> It also surfaces in more traditional R&D contracting environments: organizations seeking to procure new technology often issue requests for proposals and take bids over multiple rounds, with opportunities for feedback in between. In these settings, the sponsoring organization has better information on performance than participants or can compel interim disclosure of progress while a competition is underway.<sup>2</sup> Should it make this information known?

A similar tension is present in noninnovation organizational settings, where performance appraisals serve the dual purposes of employee development and evaluation for tournament-like promotion and retention (Beer, 1987). The observation that performance evaluation is a near-universal practice in organizations has motivated a nascent economics literature on the effects of feedback in promotion tournaments (Gershkov and Perry, 2009; Aoyagi, 2010; Ederer, 2010; Goltsman and Mukherjee, 2011), but this literature is short on empirical evidence and, with the exception of Wirtz (2014), does not account for the effects of feedback on agents' productivity. Another subliteration (Gill, 2008; Rieck, 2010; Choi, 1991; Bimpikis, Ehsani, and Mostagir, 2014; Halac, Kartik, and Liu, 2015; Mihm and Schlapp, 2016) studies disclosure policies in the context of patent races and innovation contests, but it, too, is exclusively theoretical and has excluded the possibility of feedback-driven improvement.

In this article, I use a large sample of winner-take-all logo design competitions to study the dual effects of feedback on the quantity and quality of agents' output. I first show that feedback causes players to advantageously select into continued participation and improves the quality of subsequent entries, but disclosure of intense competition discourages effort from even the top performers. A principal seeking a high-quality product thus faces a trade-off between participation and improvement in deciding whether to provide its agents with feedback. To better understand this trade-off, I estimate a structural model of the setting and use the results to simulate tournaments with alternative feedback mechanisms. The results suggest that feedback on net increases the number of high-quality designs generated, implying that it can be a valuable tool for generating innovation despite its adverse effects on incentives.

The article begins by developing a simple, illustrative model of a winner-take-all innovation contest to clarify the forces at play. In this model, a principal seeks a new product design and solicits candidates through a tournament, awarding a prize to the best entry. Players take turns submitting ideas, each of which receives immediate, public feedback revealing its quality. Partial-equilibrium predictions echo previous theoretical findings for other settings: revelation of agents' performance can be motivating for high performers, but in general it will tend to disincentivize effort by exposing leaders and laggards. However, feedback also helps agents improve at later submissions, increasing their productivity at each draw and potentially offsetting the detriment to participation. Within this framework, I characterize feedback as having two effects: a *selection effect*, which drives players with poor reviews or facing fierce competition to quit, and a *direction effect*, which guides continuing players toward increasingly better designs.

The article then transitions to an empirical study of 4294 commercial logo design tournaments from a widely used online platform. In these contests, a firm solicits custom designs from freelance designers, who compete for a winner-take-all prize awarded to the preferred entry. The contests in this sample typically offer prizes of a few hundred dollars and on average attract around 35 players and 115 designs. An essential feature of the setting is that the sponsor can

---

<sup>1</sup> Prizes have been used for centuries to encourage third parties to solve technological, commercial, and social problems. In 2010, US federal agencies received broad authority to conduct public prize competitions, with rewards ranging from status only (nonpecuniary) to tens of millions of dollars (OSTP, 2014).

<sup>2</sup> In practice, they often do. In the 2006 Netflix contest to develop an algorithm that predicts users' movie ratings, entries were immediately tested and scored, with the results posted to a public leaderboard. In the US Defense Advanced Research Projects Agency's 2005 and 2007 prize competitions to develop autonomous vehicles, participants had to publicly compete in timed qualifying races before moving on to a final round. Although these examples involve large stakes, interim scoring is also common in lower-stakes settings, such as architecture competitions or coding contests (Boudreau, Lacetera, and Lakhani, 2011; Boudreau, Lakhani, and Menietti, 2014).

provide real-time feedback on players' submissions in the form of 1- to 5-star ratings, which allow players to evaluate the quality of their own work and the competition they face. The first hints of a tension between incentives and the quality of new submissions are apparent from correlating feedback provision with contest outcomes: contests in which more designs are rated attract fewer players and designs, but are also more likely to see the sponsor award an otherwise retractable prize.

Using data at the contest-player and design level, I first provide evidence of the hypothesized effects. To identify an effect on quality, I examine (i) ratings on players' second designs, as a function of whether their first design was rated in advance; and (ii) improvements between consecutive submissions by a player when new feedback is provided but her latent performance history is unchanged. I find that feedback improves subsequent entries, especially when the previous design was poorly rated: for a player whose first design is rated 1-star, the probability that she improves with her second design goes from 26% to 51% when that rating is observed in advance (for those whose first design is rated 2-stars: 27% to 41%; for 3-stars: 17% to 26%; for 4-stars: 6% to 11%).

For evidence of an effect on incentives, I estimate the probability that a player continues participating (i) after her first rating, and (ii) at later points, as a function of ratings observed at the time. The likelihood of continuation is increasing monotonically in own ratings, with high performers continuing at a 50 percentage point higher rate than low performers following the first rating. However, even high performers can be driven away if their probability of winning is exposed to be close to zero or one. Although these effects are clearly present, the results do not resolve the extent of the trade-off between incentives and improvement, or which of the two dominates. Because this question cannot be directly answered with the data in hand, I turn to structural estimation and simulation for insight.

I develop a theoretically motivated procedure to estimate the success function and the key unobserved parameter of the model: the cost of effort. The estimated costs are relaxed to vary by player and contest and are identified from players' first-order conditions. In effect, I calculate the expected benefit of to each player's final design in a contest and the expected benefit of an additional, unentered design, and I argue that the cost must be bounded by these two quantities. Because the contests in the sample typically attract dozens of players and hundreds of designs, the difference in the gains to a player's  $n$ th design and  $(n + 1)$ th design are typically small, and the estimated bounds tight.

After validating the estimated costs against other features of the data, I use the estimates to simulate contests under policies that isolate the impacts of selection and direction. I also simulate contests in which no designs are rated (no feedback), all designs are rated (full feedback), and a random subset are rated (partial feedback, according to the frequencies observed in the data), to see how the effects of feedback vary with its frequency. As a final comparison, I simulate contests with private feedback, in which players can only see their own ratings and must infer their competitors' performance.

The simulations show that feedback reduces total participation but on net significantly increases the number of high-quality submissions. This effect grows with the frequency with which feedback is provided: partial feedback increases top-rated designs 4.5-fold, and full feedback 5.5-fold. These results suggest that private feedback might achieve the best of both worlds, allowing players to improve without exposing performance differences that discourage continued participation. The simulation of private feedback shows that this is indeed the case: despite that players expect intensifying competition, shrouding competitors' precise performance increases the total number and number of high-quality designs produced.

The simulations also reveal that direction has the dominant effect on the number of high-quality designs generated. When feedback only influences incentives (without concurrently making players more productive), the number of top-rated designs increases marginally, as a result of advantageous selection. In contrast, when improvement is enabled and selection suppressed, the number of top-rated designs explodes, increasing nearly ninefold: in effect, players

learn to make increasingly better designs, and because they grow symmetric over time and performance differences are shrouded, they continue participating longer than they otherwise would. With both channels activated, the net effect of feedback is a tempered but still large increase in the number of top-rated submissions.

Two implications follow. First, feedback can be quite valuable in competitive settings when it improves the quality or productivity of agents' effort. However, feedback that merely selects for high performers will have a limited effect on high-quality output if they cannot leverage that feedback to improve their work. The key to getting the most out of a feedback policy in this setting is thus to provide guidance while limiting attrition. The second implication is that perseverance is substantially more important to successful innovation in this setting than talent or luck: less-talented players who respond to feedback will eventually outperform more talented players who ignore this feedback or otherwise fail to improve.

The article is organized as follows. Section 1 discusses the literature in more depth and presents the theory. Section 2 introduces the empirical setting. Section 3 provides reduced-form evidence of the effects of feedback on participation and improvement. Sections 4 and 5 develop the structural model and characterize the cost estimates. Section 6 presents the results of the simulations. Section 7 concludes.

## 1. Feedback in innovation contests

□ **Existing literature on feedback in tournaments.** Rank-order tournaments have been the subject of a considerable amount of research since the seminal contributions of Tullock (1980) and Lazear and Rosen (1981), and the framework has been used to characterize competition in a wide variety of settings, most often workplace promotion. Interim evaluation in dynamic tournaments is a recent addition to the literature, motivated by the observation that “between 74% and 89% of organizations have a formal appraisal and feedback system” (Ederer, 2010; citing Murphy and Cleveland, 1995), and a battery of examples of performance evaluation in other settings.

This literature generally predicts that feedback will cause relative low performers to reduce their investment and can cause high performers to exert more or less effort. Ederer (2010) has developed a nuanced view of the problem, showing that feedback can be motivating for high performers, who learn their high productivity, but disclosure of asymmetries will discourage effort from both players, implying a trade-off between what he terms “motivation” and “evaluation” effects. Such a result follows naturally from research on the effects of asymmetries on effort in tournaments, which consistently finds that incentives of both favorites and underdogs are reduced by unbalanced competition (e.g., Baik, 1994; Brown, 2011). Though a similar sorting effect arises in the present article, the existing literature restricts attention to two-player competition. When there are many high-performing contestants, feedback may dampen incentives even if they are equally capable, as it reveals a crowded competition where the returns to effort are near zero.

The empirical evidence is both scarcer and more varied. In one example, Ederer and Fehr (2009) conduct an experiment in which agents select efforts over two periods and find that second-period effort declines in the revealed difference in first-round output. In another, Eriksson, Poulsen, and Villeval (2009) conduct an experiment in which competitors earn points for solving math problems and find that maintaining a leaderboard does not have significant effects on total attempts but can drive poor performers to make more mistakes, possibly from adopting risky strategies in trying to catch up. Azmat and Iriberry (2010) examine the effects of including relative performance information in high schoolers' report cards on their subsequent academic performance, which has a tournament-like flavor. The evidence suggests that this information has heterogeneous but on average large, positive effects on treated students' grades, which the authors attribute to increased effort.

The theoretical literature has also studied feedback or disclosure in patent races and innovation contests. Choi (1991) models a patent race and shows that information on rivals' interim

progress has offsetting effects: it exposes a discouraging technological gap, but also changes perceptions of the success rate of R&D, which can be encouraging for laggards. Bimpikis, Ehsani, and Mostagir (2014) find similar results for R&D contests. Continuing the trend, Rieck (2010) shows that enforced secrecy yields the highest expected innovation in R&D contests, as disclosure only serves to introduce asymmetry.

However, in none of these examples is feedback used by agents to improve the quality of innovation or success rates. In a setting where the goal is to realize high-quality innovation, feedback helps workers learn, reoptimize with new strategies, and ultimately improve their product. The principal focus of this part of the article is thus to bring attention to the value of feedback as guidance, as Wirtz (2014) does for organizational settings, while acknowledging the potential adverse effects on incentives.<sup>3</sup>

□ **Theoretical underpinnings.** Suppose a risk-neutral principal seeks a new product design.<sup>4</sup> Because R&D is risky and designs are difficult to objectively value, the principal cannot contract directly on performance and instead sponsors a tournament to solicit prototypes from  $J$  risk-neutral agents, who enter designs in turns. At a given turn, a player must choose whether to continue participating and if so, what idea to develop next, with each submission receiving immediate, public feedback. At the end of the tournament, the sponsor awards a winner-take-all prize  $P$  to its preferred entry. The sponsor seeks to maximize the value of the winning design.

To hone intuition, suppose each player enters at most two designs. Let each design be characterized by the following latent value  $v_{jt}$ , which only the sponsor observes:

$$v_{jt} = \ln(\beta_{jt}) + \varepsilon_{jt}, \quad \varepsilon_{jt} \sim \text{i.i.d. Type-I E.V.}, \quad (1)$$

where  $j$  indexes players and  $t$  indexes designs. In this model,  $\beta_{jt}$  represents the design's quality, which may not be known *ex ante* and is revealed by the sponsor's feedback. The design's value to the sponsor,  $v_{jt}$ , is increasing and concave in its quality, and the design with the highest  $v$  wins the contest. The  $\varepsilon_{jt}$  term is an i.i.d. random shock, which can be interpreted as idiosyncracies in the sponsor's tastes at the time a winner is chosen. Player  $j$ 's probability of winning then takes the following form:

$$Pr(\text{player } j \text{ wins}) = \frac{\beta_{j1} + \beta_{j2}}{\beta_{j1} + \beta_{j2} + \mu_j}, \quad (2)$$

where  $\mu_j \equiv \sum_{k \neq j} (\beta_{k1} + \beta_{k2})$  is player  $j$ 's cumulative competition. This success function is effectively a discrete-choice probability, obtaining directly from the primitives.

Further suppose designs are either high quality ( $\beta^H$ ) or low quality ( $\beta^L$ ). Each player enters a first design, drawn i.i.d. from a distribution  $F_\beta(\cdot)$ , yielding a high-quality design with probability  $q$  or low-quality design with probability  $1 - q$ . Players who enter a second design must decide whether to tweak their first design with minor modifications, resulting in an entry with the same  $\beta$  and a new draw of  $\varepsilon$ , or whether to draw an entirely new design from  $F_\beta(\cdot)$ , both at a common cost  $c$ . Players can abstain from a second design at no cost.<sup>5</sup>

Because  $\beta^H > E[\beta] > \beta^L$ , players who learn they have a high first draw ( $\beta^H$ ) will reuse that design, whereas those with a low first draw ( $\beta^L$ ) will redraw from  $F_\beta(\cdot)$ . Absent feedback, all players will redraw, as the expected returns to a new draw exceed that of a tweak (see Appendix A). With these behaviors in mind, the first proposition characterizes the effects of feedback on the

<sup>3</sup> As Manso (2011) shows, this type of guidance is essential to motivating innovation in single-agent settings. Wirtz (2014) offers a similar rationale for feedback in tournament settings. In both cases, the instructive effect of feedback is to inform the agent whether to keep her current technology or explore another.

<sup>4</sup> Some primitives of the model, and portions of the associated text, are borrowed from Gross (2016).

<sup>5</sup> To simplify the exposition, I assume a common  $q$  (probability of a high-quality draw) and  $c$  (cost), though these parameters could be relaxed to vary across players to accommodate *ex ante* differences in ability or costs—the results of this section do not depend on their precise values.

quality of a player's second design, which I term a *direction* effect: feedback guides players toward higher-quality subsequent draws.

*Proposition 1.* Feedback yields greater quality improvements over a player's initial entry, with low performers in expectation improving more than high performers. In notation:

- (a)  $E[\beta_{j2} - \beta_{j1} | \text{Feedback}] > E[\beta_{j2} - \beta_{j1} | \text{No feedback}]$ ,  
 (b) Conditional on feedback,  $E[\beta_{j2} - \beta_{j1} | \beta_{j1} = \beta^L] > E[\beta_{j2} - \beta_{j1} | \beta_{j1} = \beta^H]$ .

The rest of this section provides two illustrative results in partial equilibrium which demonstrate feedback's *selection* effects. First, observe that in the absence of feedback, players are strategically symmetric, with equal probabilities of having a high- or low-initial draw, whereas feedback introduces the possibility of asymmetries. Proposition 2 shows that these asymmetries reduce players' incentives to enter a second design, resembling Ederer's (2010) evaluation effect. Proposition 3 then shows that provided competition is sufficiently high, as it typically is in the data, players with better feedback have greater incentives to continue participating than those with worse feedback, similar to Ederer's (2010) motivation effect. Proofs of all propositions are provided in Appendix A.

*Proposition 2.* The returns to a player's second design decline to zero as the quality of her first design  $\beta_{j1}$  and the cumulative competition  $\mu_j$  grow distant.

*Proposition 3.* There exists a level of competition  $\mu^*$  such that for all  $\mu_j > \mu^*$ ,

$$\underbrace{\left( \frac{2\beta^H}{2\beta^H + \mu_j} \right) - \left( \frac{\beta^H}{\beta^H + \mu_j} \right)}_{\text{Benefit to effort|pos. feedback}} > q \underbrace{\left( \frac{\beta^L + \beta^H}{\beta^L + \beta^H + \mu_j} \right) + (1 - q) \left( \frac{2\beta^L}{2\beta^L + \mu_j} \right) - \left( \frac{\beta^L}{\beta^L + \mu_j} \right)}_{\text{Benefit to effort|neg. feedback}}.$$

Intuitively, we might expect that negative feedback will induce quitting: players with poor feedback are not only fighting an uphill battle, but they are also more likely to produce lower-quality ideas. Conversely, distant favorites can be equally unmotivated to exert effort, as victory is nearly assured—though this scenario is less likely to occur in a large field of competitors. In settings with many players, an additional result emerges: high performers who face heavy competition will also have low incentives for effort, because competition flattens the payoffs to the point where marginal returns fall below cost.

Though these results abstract from the effects of strategic interactions, more explicit consideration would only tend to reinforce them. Effort from leaders would then yield indirect benefits by discouraging followers, similar to the Fudenberg et al. (1983) notion of  $\varepsilon$ -preemption in the context of patent races, but this benefit would dissipate when competition is very high and future efforts already deterred.

## 2. Graphic design contests

■ I use a sample of 4294 commercial logo design contests from a popular online platform to study the effects of feedback in a competitive product development setting. The platform hosts hundreds of contests each week in several categories of design, with logos being the modal category and thus a natural choice for analysis. Note that a firm's choice of logo is also nontrivial, as it is the defining feature of its brand, which can be one of its most valuable assets, and is how consumers will identify it for years to come.

In these contests, a firm (the sponsor, typically a small business) solicits custom designs from freelance designers (players) in exchange for a fixed prize awarded to its favorite entry. The sponsor publishes a design brief describing its business, its customers, and what it seeks to communicate with its logo; specifies the prize structure; sets a deadline for entries; and opens

the contest to competition. While the contest is open, players can enter as many designs as they want, at any time they want, and sponsors can provide players with real-time feedback on their submissions in the form of 1- to 5-star ratings and written commentary. Players can see competing designs and the distribution of ratings on these designs, but not the ratings on specific competing designs. Copyright is enforced. At the end of the contest, the sponsor picks the winning design and receives the design files and full rights to their use. The platform then transfers payment to the winner.

Appendix B describes the data set in detail. For each contest in the sample, I observe the design brief, the beginning and ending dates, the prize amount, and whether the prize is committed. Though multiple prizes are possible, the sample is restricted to contests with a single, winner-take-all prize. I further observe every submission, the identity of the designer, his or her history on the platform, the time and order in which the design was entered, the rating it received (if any), the time at which the rating was given, and whether it won the contest. I also observe when players withdraw designs from the competition, but I assume withdrawn entries remain in contention, as sponsors can request that any withdrawn design be reinstated. Because I do not observe written feedback for all contests, I assume the content of written commentary is fully summarized by the rating.<sup>6</sup>

The player identifiers allow me to track players' activity over the course of each contest and across contests in other design categories dating back to the platform's creation. I use the precise timing information to reconstruct the state of a contest at the time each design is submitted. For every design, I calculate the number of preceding designs in the contest of each rating. I do so both in terms of the prior feedback available (observed) at the time of submission as well as the feedback eventually provided. To account for the lags required to produce a design, I define preceding designs to be those entered at least one hour prior to a given design and similarly require that feedback be provided at least one hour prior to the given design's submission to be considered observed when it was made.

□ **Characteristics of the sample.** The average contest in the sample lasts nine days, offers a \$295 prize, and attracts 116 designs from 37 players; on average, 56% of these are rated (Table 1). A contest will typically generate just over three 5-star designs. By default, the sponsor retains the option of not awarding the prize to any design if none are to its liking, but the sponsor can forgo this option and commit to awarding the prize when it creates the contest. Though only 23% of contests have a committed prize, 89% are awarded.

The median contest entrant submits two designs and receives one rating. The median player entered two contests within this sample and produced six designs, though these distributions are heavily skewed, with many players entering only one or two contests and a few participating frequently over extended periods. Only 10% (1488) of these players win a contest in the sample, implying the presence of repeat winners.

Table 2 shows the distribution of ratings. A total of 58% of the designs in the sample (285,082 of 496,041) are rated, with a median and modal rating of 3 stars. Only 5% of rated designs in the data receive the top, 5-star rating, suggesting that sponsors reserve this category for their most preferred entries. Indeed, as Appendix Table D.2 shows, a disproportionate number (almost 40%) of winning designs are rated 5 stars, and nearly 75% are rated 4 or more stars. Although the ratings convey a substantial amount of information about a design's odds of success, they do not perfectly predict them: 1- and 2-star designs occasionally win contests—albeit with very low frequency (0.4% and 1.0% of awarded contests in the sample, respectively)—suggesting that an element of luck remains until the end. Explanations for why low-rated designs sometimes win, or more generally, why 5-star designs do not always win, include last-minute changes of heart

<sup>6</sup> Written feedback is observed for a small subset of the sample. Where observed, comments are given only to a small fraction of designs in a contest (on average, 12%), far less than are rated, and typically echo the rating given, suggesting they do not play an important role in this setting.

**TABLE 1** Descriptive Statistics

Variable	N	Mean	Standard Deviation	P25	P50	P75
<b>Panel A. Characteristics of Contests in the Sample</b>						
Contest length (days)	4294	9.15	3.72	7	7	13
Prize value (US\$)	4294	295.22	128.12	200	250	350
No. of players	4294	37.28	25.35	23	31	43
No. of designs	4294	115.52	94.82	65	92	134
5-star designs	4294	3.41	6.97	0	1	4
4-star designs	4294	13.84	17.89	3	9	19
3-star designs	4294	22.16	26.99	5	15	30
2-star designs	4294	16.04	23.36	2	8	21
1-star designs	4294	10.94	28.78	0	3	12
Unrated designs	4294	49.14	63.36	10	34	65
Number rated	4294	66.38	73.34	21	50	88
Fraction rated	4294	0.56	0.33	0.3	0.6	0.9
Prize committed	4294	0.23	0.42	0.0	0.0	0.0
Prize awarded	4294	0.89	0.31	1.0	1.0	1.0
<b>Panel B. Characteristics of Contest-Players in the Sample</b>						
No. of designs	160,059	3.10	3.53	1	2	4
Number rated	160,059	1.78	2.95	0	1	2
<b>Panel C. Characteristics of Players in the Sample</b>						
Total contests	14,843	10.78	32.75	1	2	7
Total submissions	14,843	33.42	107.58	1	6	21
Ever a winner?	14,843	0.10	0.30	0	0	0

Notes: Panel A reports descriptive statistics for the sampled contests. “Fraction rated” refers to the fraction of designs in each contest that get rated. “Prize committed” indicates whether the contest prize is committed to be paid (vs. retractable). “Prize awarded” indicates whether the prize was awarded. The fraction of contests awarded subsumes the fraction committed, as committed prizes are always awarded. Panel B reports descriptives at the level of contest-players (one observation per player per contest), and Panel C at the level of unique players.

**TABLE 2** Distribution of ratings (rated designs only)

	1-Star	2-Star	3-Star	4-Star	5-Star	Total
<b>Count</b>	46,983	68,875	95,159	59,412	14,623	<b>285,052</b>
<b>Percent</b>	16.48	24.16	33.38	20.84	5.13	<b>100</b>

Notes: Table tabulates rated designs by rating. 57.5% of designs in the sample are rated by sponsors on a 1 to 5 scale. The site provides guidance on the meaning of each rating, which introduces consistency in the interpretation of ratings across contests.

or differences of opinion between the member of the sponsoring organization administering the ratings and the person or committee selecting the winner.

□ **Correlations of contest characteristics with outcomes.** To shed light on how the sampled contests operate and how different levers affect outcomes, Table 3 relates contest-level outcomes with contest characteristics. The specifications in columns (1) to (3) regress the number of players, designs, and designs per player on the prize value, contest duration, length of the design brief, number of materials provided to be included in the design, and fraction of designs rated, as well as the average cost of participating players (estimated in later sections). Most of these variables are fixed by the sponsor before the contest begins, and although the fraction of entries rated and players’ costs are in part endogenously determined during the contest, in practice they largely reflect the sponsor’s type (engaged or aloof) and the difficulty of the project.

TABLE 3 Correlations of Contest Outcomes with Their Characteristics

	(1) Players	(2) Designs	(3) Designs/Player	(4) Awarded
Total prize value (\$100s)	14.741*** (0.659)	54.723*** (2.524)	0.116*** (0.015)	0.234*** (0.041)
Committed value (\$100s)	1.856* (1.115)	5.652 (4.399)	0.010 (0.025)	
Average cost (\$)	-2.637*** (0.129)	-12.989*** (0.497)	-0.122*** (0.006)	-0.192*** (0.014)
Fraction rated	-13.618*** (0.783)	-15.819*** (2.791)	0.733*** (0.040)	0.753*** (0.105)
Contest length	0.340*** (0.068)	1.137*** (0.251)	0.004 (0.004)	0.008 (0.010)
Words in desc. (100s)	0.076 (0.080)	2.962*** (0.389)	0.060*** (0.005)	-0.157*** (0.014)
Attached materials	-0.867*** (0.160)	-1.510** (0.604)	0.051*** (0.012)	-0.010 (0.016)
Prize committed	1.104 (3.282)	3.091 (12.907)	-0.021 (0.085)	
Constant	8.942*** (1.746)	-7.648 (6.266)	2.445*** (0.074)	1.947*** (0.178)
<i>N</i>	4294	4294	4294	3298
<i>R</i> <sup>2</sup>	0.63	0.65	0.30	

Notes: Table shows the estimated effect of contest attributes on overall participation and the probability that the prize is awarded. The final specification is estimated as a probit on contests without a committed prize. \*, \*\*, \*\*\* represent significance at the 0.1, 0.05, and 0.01 levels, respectively. Monthly fixed effects included but not shown. Robust standard errors in parentheses.

An additional \$100 in prize value on average attracts around 15 more players and 55 more designs. The effects of feedback can be similarly powerful: relative to a sponsor who rates no designs, one who rates every design will typically attract 14 fewer players and 16 fewer designs. Other features have more modest effects.

Column (4) models the probability that a sponsor awards an uncommitted prize, implying that the contest produced a design good enough *to be* awarded. Feedback dramatically increases this probability, suggesting that feedback is critical to the development of high-quality work. In light of the aforementioned evidence that feedback reduces participation, this result provides the first indication of a tension between attracting *more* effort versus *higher-quality* effort. Contests with larger prizes, shorter descriptions, and lower costs are also more likely to be awarded, though the size of these effects is considerably smaller than the effects of feedback: the impact of full feedback (relative to no feedback) on the probability the prize is awarded is nearly equal to that of a \$300 prize increase—a more than doubling of the average and median prize in the sample.

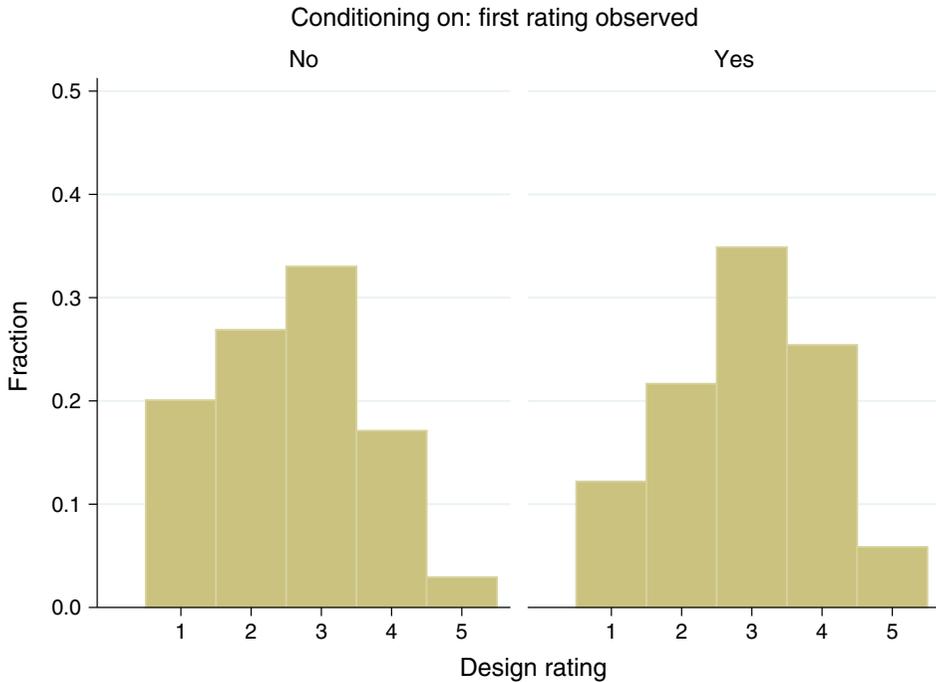
### 3. Reduced-form evidence of selection and direction

□ **Effects of feedback on quality.** In evaluating the effects of feedback on the quality of new submissions, a natural starting point is to examine the distribution of ratings on a player's second design, conditional on whether her first design was rated before the second was entered. As Figure 1 shows, the overall mass of the ratings distribution shifts upward when feedback is observed in advance—although this pattern could potentially be confounded by selection if sponsors are more likely to give timely feedback to the high-performing players.

To eliminate this concern, I condition the comparison on the first design's rating. Figure 2 shows the distribution of ratings on a player's second design, conditional on the rating that has or will be given on her first design and whether observed in advance.

FIGURE 1

DISTRIBUTION OF RATINGS ON SECOND DESIGN, BY WHETHER FIRST DESIGN'S RATING OBSERVED IN ADVANCE [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]



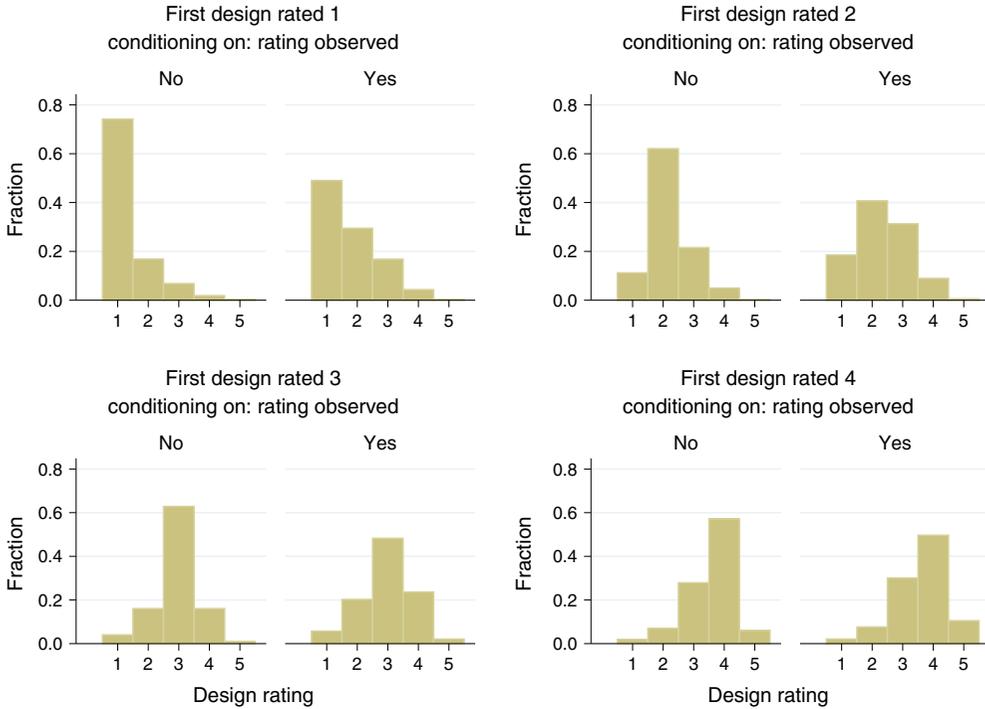
Notes: Figure reports the distribution of ratings on a player's second design in a contest conditional on whether that player's first design was rated prior to making the second one. The fraction rated 5 in the left panel is 0.03. The fraction rated 5 in the right panel is 0.06. The difference is significant with  $p=0.00$ .

Table 4 provides the accompanying differences in means and shows that players improve at significantly higher rates when they observe feedback in advance. Among players with a 1-star design, 51% score a higher rating on their second entry when they observe feedback in advance, versus 26% among those who do not (Panel A). For players with a 2-star design, the percentages are 41% and 27%; for players with a 3-star design, 26% and 17%; and for players with a 4-star design, 11% and 6%, with all differences precisely estimated and significant at the 1% level. Panel B shows the estimated effect on the second design's rating in levels, conditional on weak improvement.

These comparisons could nevertheless still be confounded if the players who are most likely to *improve* disproportionately wait for feedback before entering a second design. To address this concern, Appendix C provides robustness checks examining pairs of consecutive submissions by a given player, in a given contest, between which the player's information set may change due to feedback but her latent ratings history remains the same. As in the results for initial ratings, I find that newly-arrived ratings equal to or higher than the rating on the previous design results in a greater likelihood of improvement, whereas revelation of lower ratings has a precisely-estimated zero effect. Absent new information, the next submission tends to be of roughly the same quality as the prior submission, but feedback has large effects: players who enter 1- or 2-star designs and then receive a 5-star rating on an earlier submission tend to improve by over a full point more than they otherwise would. The magnitude of these effects expectedly declines the better the prior submission, reflecting the difficulty of improving from a high initial value.

FIGURE 2

DISTRIBUTION OF RATINGS ON SECOND DESIGN, BY RATING ON FIRST DESIGN AND WHETHER OBSERVED IN ADVANCE [Color figure can be viewed at wileyonlinelibrary.com]



Notes: Figure shows the distribution of ratings on a player's second design in a contest, conditioning on (i) the rating they receive on their first design, and (ii) whether that rating was observed prior to entering their second design. In all subfigures, the fraction of players who score a better rating on their second design than on their first is higher when they observe feedback in advance (for players with a 1-star: 51% versus 26%; for those with a 2-star: 41% versus 27%; for those with a 3-star: 26% versus 17%; for those with a 4-star: 11% versus 6%). All differences are significant at the 1% level.

□ **Effects of feedback on participation.** I again look to the first rating a player receives to identify the effects of feedback on effort choices. I focus on a player's first rating because it is typically the first indication of the sponsor's preferences and thus *ex-ante* unpredictable, and because the median player enters only two designs and receives one rating. Appendix C nevertheless reports estimates from an analogous examination of second ratings, with similar results.

Figure 3 provides descriptive evidence of these effects. The figure shows the distribution of the number of designs a player enters after receiving her first rating, conditional on whether that first rating is 1-star (left-most panel), or 4- or 5-stars (right-most panel). A majority (69.5%) of players whose first rating is the lowest possible rating will subsequently stop investing in that contest. In contrast, the majority (61.2%) of players who receive a high rating will subsequently enter at least one more design.

To formalize this evidence, I estimate the effect of each player's first rating and the competition observed at the time it is made on the probability that the player subsequently abandons the contest, projecting an indicator for abandonment on indicators for each rating she may have received as well as indicators for the highest rating on competing designs. Table 5 estimates variants of the following specification:

$$Abandon_{jk} = \beta_0 + \sum_{r=1}^5 \beta_r \cdot \mathbb{1}(R_{jk} = r) + \sum_{r=1}^5 \gamma_r \cdot \mathbb{1}(\bar{R}_{-jk} = r)$$

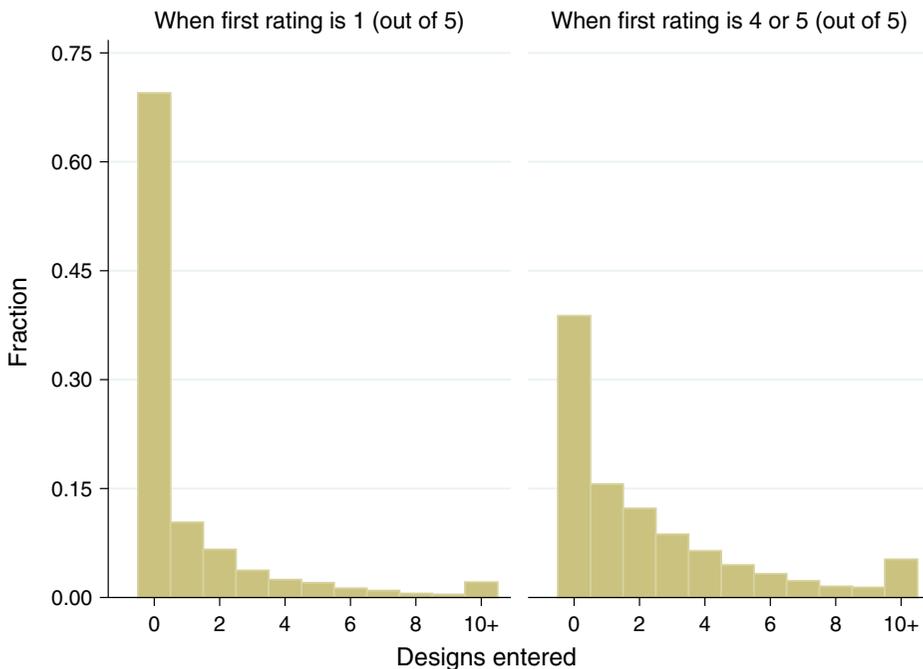
**TABLE 4** Effects of feedback: Improvement between First and Second Submissions

	When First Design Rated:			
	1-Star	2-Star	3-Star	4-Star
<b>Panel A. Probability of improvement</b>				
$\mathbb{1}(\text{First rating observed})$	0.251*** (0.015)	0.140*** (0.010)	0.087*** (0.007)	0.045*** (0.006)
Constant	0.258*** (0.006)	0.267*** (0.005)	0.171*** (0.004)	0.060*** (0.003)
<i>N</i>	8466	12653	16739	9192
<i>R</i> <sup>2</sup>	0.04	0.02	0.01	0.01
<b>Panel B. Rating of Second Design</b>				
$\mathbb{1}(\text{First rating observed})$	0.403*** (0.026)	0.260*** (0.016)	0.150*** (0.010)	0.079*** (0.010)
Constant	1.373*** (0.011)	2.362*** (0.007)	3.226*** (0.005)	4.095*** (0.005)
<i>N</i>	8466	11017	13072	5701
<i>R</i> <sup>2</sup>	0.04	0.03	0.02	0.01

Notes: Table shows the effects of observing feedback in advance of a player's second design in a contest on the probability that it is higher rated than her first entry (Panel A) and on its rating, conditional on weakly improving (Panel B). \*, \*\*, \*\*\* represent significance at the 0.1, 0.05, and 0.01 levels, respectively. Standard errors clustered by player in parentheses.

**FIGURE 3**

DESIGNS ENTERED AFTER A PLAYER'S FIRST RATING [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]



Notes: Figure shows the distribution number of designs entered by players after receiving their first rating in a contest, by the value of that first rating. A total of 86,987 contest-players received first ratings. Of these: 16,374 were rated 1-star (18.8%); 22,596 were rated 2-stars (26.0%); 28,945 were rated 3-stars (33.3%); 16,233 were rated 4-star (18.7%); and 2839 were rated 5-star (3.3%). The figure illustrates that players are much more likely to continue participating in a contest after positive feedback.

**TABLE 5** Tendency to Abandon After First rating, as Function of Rating

	Dependent variable: abandon after first rating			
	(1) Linear	(2) Linear	(3) Linear	(4) Logit
Player's first rating==5	-0.444*** (0.017)	-0.394*** (0.017)	-0.485*** (0.020)	-2.311*** (0.086)
Player's first rating==4	-0.437*** (0.010)	-0.385*** (0.010)	-0.454*** (0.012)	-2.269*** (0.055)
Player's first rating==3	-0.280*** (0.008)	-0.242*** (0.008)	-0.290*** (0.009)	-1.468*** (0.044)
Player's first rating==2	-0.114*** (0.007)	-0.097*** (0.007)	-0.120*** (0.008)	-0.620*** (0.038)
Competitors' prior best==5	0.037*** (0.013)	0.064*** (0.011)	0.058*** (0.014)	0.203*** (0.065)
Constant	0.483*** (0.026)	0.465*** (0.020)	0.174 (0.106)	-0.898 (1.130)
<i>N</i>	48125	48125	48125	46935
<i>R</i> <sup>2</sup>	0.24	0.36	0.46	
Contest FEs	Yes	No	Yes	Yes
Player FEs	No	Yes	Yes	No

Notes: Table shows the effect of a player's first rating in a contest and the competition at that time on the probability that the player subsequently enters no more designs. Observations are contest-players. The dependent variable in all columns is an indicator for whether the player abandons after her first rating. Columns (1) to (3) estimate linear models with fixed effects; Column (4) estimates a logit model without player fixed effects, which may render the estimates inconsistent. \*, \*\*, \*\*\* represent significance at the 0.1, 0.05, and 0.01 levels, respectively. All specifications control for time remaining, both in levels and as a percent of the contest duration. Standard errors clustered by player in parentheses.

$$+ \delta \cdot \text{Timing}_{jk} + X_{jk}\theta + \zeta_k + \varphi_j + \varepsilon_{jk},$$

where  $Abandon_{jk}$  indicates that player  $j$  entered no designs in contest  $k$  after her first rating;  $R_{jk}$  is the player's first rating;  $\bar{R}_{ijk}$  is the highest rating on any competing designs at that time;  $Timing_{jk}$  measures the fraction of the contest elapsed at the time of that first rating;  $X_{jk}$  is a vector of controls; and  $\zeta_k$  and  $\varphi_j$  are contest and player fixed effects, respectively. Although it is likely the case that players with no observed activity after their first rating made a deliberate choice to stop investing, this measure cannot distinguish players who immediately abandon from those who "wait and see" and later abandon, though the distinction is immaterial for the purposes of this exercise.

Columns (1) to (3) estimate linear probability models with contest, player, and contest and player fixed effects, respectively. Linear models are used in order to control for these fixed effects (especially player fixed effects), which may not be estimated consistently in practice and could thus render the remaining estimates inconsistent in a binary outcome model. Column (4) estimates a logit model with only contest fixed effects. The linear model with two-way fixed effects (Column 3) is the preferred specification.

The probability that a player continues investing is monotonically increasing in her first rating. The magnitudes are large: players with a 5-star first rating remain active at a precisely estimated 50 percentage point higher rate than those with a 1-star first rating. High-rated competition also makes it more likely that a player abandons after her first rating. Together, these results demonstrate that feedback generates advantageous selection, with the high-rated players more likely to continue competing and low performers opting out after receiving low marks, but by revealing high rated competition, feedback can have the perverse consequence of driving highly rated players away.

To bring the effects of asymmetry into focus, I estimate a similar model, replacing the indicators with a quadratic in a player's probability of winning upon receiving her first rating,

**TABLE 6** Tendency to Abandon After First Rating, as Function of Pr(Win)

	Dependent variable: abandon after first rating			
	(1) Linear	(2) Linear	(3) Linear	(4) Logit
Pr(Win)	-1.643*** (0.049)	-1.565*** (0.053)	-1.644*** (0.055)	-8.756*** (0.280)
Pr(Win) <sup>2</sup>	1.565*** (0.055)	1.478*** (0.061)	1.537*** (0.061)	8.284*** (0.306)
Constant	0.398*** (0.025)	0.379*** (0.019)	-0.052 (0.085)	-1.622 (1.389)
<i>N</i>	48125	48125	48125	46935
<i>R</i> <sup>2</sup>	0.20	0.34	0.43	
Contest FEs	Yes	No	Yes	Yes
Player FEs	No	Yes	Yes	No
Minimizer	0.52	0.53	0.53	0.53

Notes: Table shows the effect of a player's win probability after receiving her first rating on the probability that she subsequently enters no more designs. Observations are contest-players. The dependent variable in all columns is an indicator for whether the player abandons after her first rating. Columns (1) to (3) estimate linear models with fixed effects; Column (4) estimates a logit model without player fixed effects, which may render the estimates inconsistent. \*, \*\*, \*\*\* represent significance at the 0.1, 0.05, and 0.01 levels, respectively. All specifications control for time remaining, both in levels and as a percent of the contest duration. Standard errors clustered by player in parentheses.

which can be computed from the conditional logit model estimated later in the article (Section 4). Table 6 shows results from a similar arrangement of specifications: linear models with contest, player, and contest and player fixed effects in Columns (1) to (3), and a logit model with contest fixed effects in Column (4).

The tendency to abandon is definitively convex in a player's probability of winning, reaching a minimum near a win probability of 0.5, and a maximum at the outer bounds of a win probability of 0 and 1. The estimates are statistically similar across all specifications and concord with theoretical predictions that incentives for effort are greatest when agents are running even with their competition (e.g., Baik, 1994).

□ **Results in context.** The collective evidence shows that feedback has the desirable effect of improving the quality of future submissions and nudging poor performers out of contention, reducing wasteful effort and concentrating incentives for the remaining participants. However, feedback can simultaneously reduce incentives for high performers to participate, relative to incentives in a state of ignorance, by revealing or enabling high-quality competitors.

The principal thus faces a fundamental trade-off between participation and improvement. Given that sponsors who provide the most feedback are the most likely to award a retractable prize (recall Table 3), it would seem that feedback has a large enough effect on quality to be desirable. Yet the reduced form is ultimately inconclusive and cannot discern how much of the increase in quality is due to better players or better designs. The distinction is not only important for understanding how to deliver feedback, but also revealing of the relative contributions of ability versus improvement to success in this setting.

If selection or direction were constrained, or if feedback were provided at varying frequency or made private, what would be the effects on total and high-quality output? To answer these questions, I develop a procedure to estimate effort costs and use the estimates to simulate and evaluate alternative feedback mechanisms.

## 4. Structural model

■ The empirical model borrows ideas from the empirical auctions literature, which uses theoretical insights to estimate unobserved distributions of bidder values, and it is flexible in

that costs are allowed to vary by contest and player, reflecting the fact that some contests are more demanding than others and that players have heterogeneous reservation wages. The main assumptions are (i) that each player has a constant cost in a given contest, and (ii) that players compete until this cost exceeds the expected benefit. With a consistently estimated success function, a given player's cost in a given contest will be set-identified in a sample with *any* number of contests or players. The bounds of the identified set converge on a point as the number of players or designs in a contest grows large (as they do in this article), irrespective of the number of contests in the sample.

The estimation proceeds in two steps. In the first step, I estimate a logistic success function that translates players' effort into their probability of winning. I then combine the success function with nonparametric frequencies of ratings on a player's next design, conditional on her prior history in the contest, to calculate the expected payoff to each player's last design in a contest and the "extra" design that the player chose not to enter. Under the assumption that the game ends in a complete information Nash equilibrium, these quantities place bounds on cost: a player's cost must be less than the expected benefit from her final design but greater than the expected benefit of an additional, unentered design. The logic behind this procedure is closely related to that of Haile and Tamer (2003), who use an analogous approach with drop-out bid levels to put bounds on the distribution of latent bidder values in symmetric English auctions.

Denote the rating of design  $i$  from player  $j$  in contest  $k$  as  $R_{ijk}$ , and let  $R_{ijk} = \emptyset$  when the sponsor declines to rate  $ijk$ . Ratings provide players with information on two unknowns: (i) the likelihood of a given design winning the contest, conditional on the competition, and (ii) how well her next design is likely to be rated, should she decide to make one. To make the intuition concrete, consider the following cases:

- (a)  $R_{ijk} = 5$ . The design is outstanding and has a very high chance of winning the contest. The player has caught onto a theme that the sponsor likes. The player's subsequent designs are likely to be highly rated as well. (Though note that the marginal increase to the player's odds of winning the contest from another submission fall dramatically, because she already entered a design that is a strong contender. Any other 5-star designs she enters will substantially cannibalize  $ijk$ 's odds.)
- (b)  $R_{ijk} = 1$ . The design is not good and is unlikely to win. The player hasn't yet figured out what the sponsor likes, and her next design will likely be poorly rated as well.
- (c)  $R_{ijk} = \emptyset$ . The player receives no feedback on her design. She has no new information, and the distribution of ratings on her next design is roughly unchanged.

In the model below, I formalize this intuition. The empirical model treats the design process as a series of experiments that adapts to feedback from the sponsor, as in the theoretical model of Section 1 and in Gross (2016). Players use feedback to determine the probability that the rated design wins the contest, refine their experimentation, and set expectations over the ratings on any subsequent designs. Using the nonparametric distribution of ratings on a design in conjunction with a conditional logit model translating those outcomes into contest success, players then (i) calculate the expected benefit of another design, (ii) compare it to their contest-, player-specific cost, and (iii) participate until the costs exceed the benefits. The cost of design is set-identified from this stopping choice. I take the midpoint of the set as a point estimate.

The estimated cost will be the cost of making a single design and will be that which rationalizes the stopping choices observed in the data. I assume that this cost is constant for each player throughout a given contest. Design costs primarily reflect the opportunity cost of the time and resources a player expends in the activity of designing a logo. They thus reflect differences in reservation wages, which figure prominently when players can enter from anywhere in the world. However, the estimated costs may also reflect (and net out) any unobserved payoffs in the form of learning, practice, and portfolio-building, all of which motivate players' participation. There may also be an unobserved payoff that accrues to the winner, such as a new client relationship; the expected value of this benefit will be captured in the estimates as well. Finally, the estimates

will also reflect any level bias that a player has over the process determining her probability of winning. In effect, I will be measuring the costs that players behave as if they face.

□ **Details: framework and estimation.**

*Estimating the success function.* Let  $i$  index submissions,  $j$  index players, and  $k$  index contests. Suppose every contest  $k$  has  $J_k > 0$  risk-neutral players, each of whom makes  $I_{jk} > 0$  submissions. Let  $I_k = \sum_{j \in J_k} I_{jk}$  be the total number of designs in contest  $k$ . Players in contest  $k$  compete for a prize  $P_k$ .

As in the theoretical model, I assume the sponsor awards the prize to its preferred design. Formally, let  $v_{ijk}$  be the latent value of design  $ijk$  to the sponsor of contest  $k$ , and suppose that this value is a function of the design's rating and an i.i.d. Type-I E.V. error. With six possibilities for ratings, a design's value can be written parsimoniously as the sum of fixed effects for each rating and an error term:

$$v_{ijk} = \gamma_0 \mathbb{1}(R_{ijk} = \emptyset) + \gamma_1 \mathbb{1}(R_{ijk} = 1) + \dots + \gamma_5 \mathbb{1}(R_{ijk} = 5) + \varepsilon_{ijk} \equiv \psi_{ijk} + \varepsilon_{ijk}. \quad (3)$$

This specification is closely related to the theoretical model in equation (1), with the main difference being a restricted, discrete domain for quality. The error term represents unpredictable variation in the sponsor's preferences and explains why 5-star designs do not always win. Although the number and content of designs in the sponsor's choice set varies between contests, designs always share a common attribute in their rating, which is assumed to fully characterize the predictable component of a design's quality, including any written commentary provided to players but not observed in the data set. The choice set is assumed to satisfy I.I.A.; in principle, adding a design of any rating to a given contest would reduce competing designs' chances of winning proportionally.<sup>7</sup> For contests with an uncommitted prize, the choice set includes an outside option of not awarding the prize, whose value is normalized to zero. Under this model, player  $j$ 's probability of winning is:

$$Pr(j \text{ wins } k) = \frac{\sum_{i \in I_{jk}} e^{\psi_{ijk}}}{\sum_{i \in I_k} e^{\psi_{ik}} + \mathbb{1}(\text{Uncommitted prize})}.$$

This success function can be estimated as a conditional logit model (McFadden, 1974) using the rating and win-lose outcome of every design in the sample. The estimates are provided in Appendix Table D.1, from which several patterns emerge. First, the value of a design is monotonically increasing in its rating, with only a 5-star rating on average preferred to the outside option, and the fixed effects are precisely estimated. To produce the same change in the success function generated by a 5-star design, a player would need 12 4-star designs, 137 3-star designs, or nearly 2000 1-star designs. As a measure of fit, the predicted odds-on favorite wins almost half of all contests in the sample. These results demonstrate that this simple model fits the data quite well and in an intuitive way, suggesting that ratings provide considerable information about a player's probability of winning.

*Calculating the expected benefit from a design.* To compute the expected benefit to a given player of entering an additional design, I consider all of the ratings it may receive (ratings of 1 to 5, or no rating), calculate the incremental change in the success function under each possibility, and take the weighted average, weighting by the non parametric probability of obtaining each rating conditional on a player's history in the same contest, which flexibly incorporates learning.

Let  $s_{ijk}$  be a state variable characterizing the *eventual* ratings on all of player  $j$ 's designs in contest  $k$  made *prior* to her  $i^{\text{th}}$  design. In other words, it's the true quality of all of  $j$ 's submissions

<sup>7</sup> I also test the I.I.A. assumption by removing subsets of designs from each contest and reestimating the model. The results are statistically and quantitatively similar when the choice set is deliberately varied.

prior to her  $i$ th one, to the best anyone will ever know.  $s_{ijk}$  can be summarized with the following six-dimensional vector:

$$s_{ijk} = \left[ \sum_{x < i} \mathbb{1}(R_{xjk} = \emptyset), \sum_{x < i} \mathbb{1}(R_{xjk} = 1), \dots, \sum_{x < i} \mathbb{1}(R_{xjk} = 5) \right].$$

In practice, a player’s earlier designs aren’t always rated before she makes her next one. Because  $s_{ijk}$  incorporates all information on  $j$ ’s past designs that will ever be known, we can think of it as the ratings history under omniscience. However, players’ design choices must be made on the basis of prior, *observed* ratings. Let  $\tilde{s}_{ijk}$  be the ratings on previous submissions that player  $j$  observes at the time of her  $i$ th submission in contest  $k$ . Writing  $\tilde{R}_{xjk}$ ,  $x < i$  as the rating on submission  $x < i$  observed by a player at the time of her  $i$ th submission, we can write the observable ratings history as:

$$\tilde{s}_{ijk} = \left[ \sum_{x < i} \mathbb{1}(\tilde{R}_{xjk} = \emptyset), \sum_{x < i} \mathbb{1}(\tilde{R}_{xjk} = 1), \dots, \sum_{x < i} \mathbb{1}(\tilde{R}_{xjk} = 5) \right].$$

The sum of the entries in the vector  $\tilde{s}_{ijk}$  will equal the sum of those in  $s_{ijk}$ , but in general the null rating count will be higher for  $\tilde{s}_{ijk}$ , as some submissions made prior to  $i$  that will eventually be rated will not yet have been rated at the time of the  $i$ th submission.

With a sample of 496,041 submissions, the nonparametric distribution of the rating on  $j$ ’s  $i$ th submission, conditional on her observable ratings history,  $\tilde{s}_{ijk}$  can be estimated as:

$$\hat{f}(R_{ijk} = r | \tilde{s}_{ijk}) = \frac{\sum_{\ell \in I_k, k \in K} \mathbb{1}(R_{\ell k} = r | \tilde{s}_{\ell k} = \tilde{s}_{ijk})}{\sum_{\ell \in I_k, k \in K} \mathbb{1}(\tilde{s}_{\ell k} = \tilde{s}_{ijk})} \xrightarrow{d} f(r | \tilde{s}_{ijk}).$$

In words, the probability that player  $j$ ’s  $i$ th design in contest  $k$  is rated  $r$ , given an observable ratings history of  $\tilde{s}_{ijk}$ , can be estimated from the data as the fraction of all designs in the data made in state  $\tilde{s}_{ijk}$  that received the rating  $r$ . With a small, discrete sample space, these probabilities can be easily estimated without the kernel methods required for nonparametric estimation of continuous distributions.

The distribution  $\hat{f}(\cdot)$  nevertheless suffers a curse of dimensionality due to the large heterogeneity in ratings histories. To reduce the dimensionality, I redefine  $s_{ijk}$  and  $\tilde{s}_{ijk}$  to indicate *whether* a player has received each rating, as opposed to counts of each rating. This adjustment is designed to make the nonparametric estimation tractable (with  $2^6 = 64$  cells) while retaining the most important information in the ratings history. Under this construction,  $\tilde{s}_{ijk}$  can be redefined as follows:

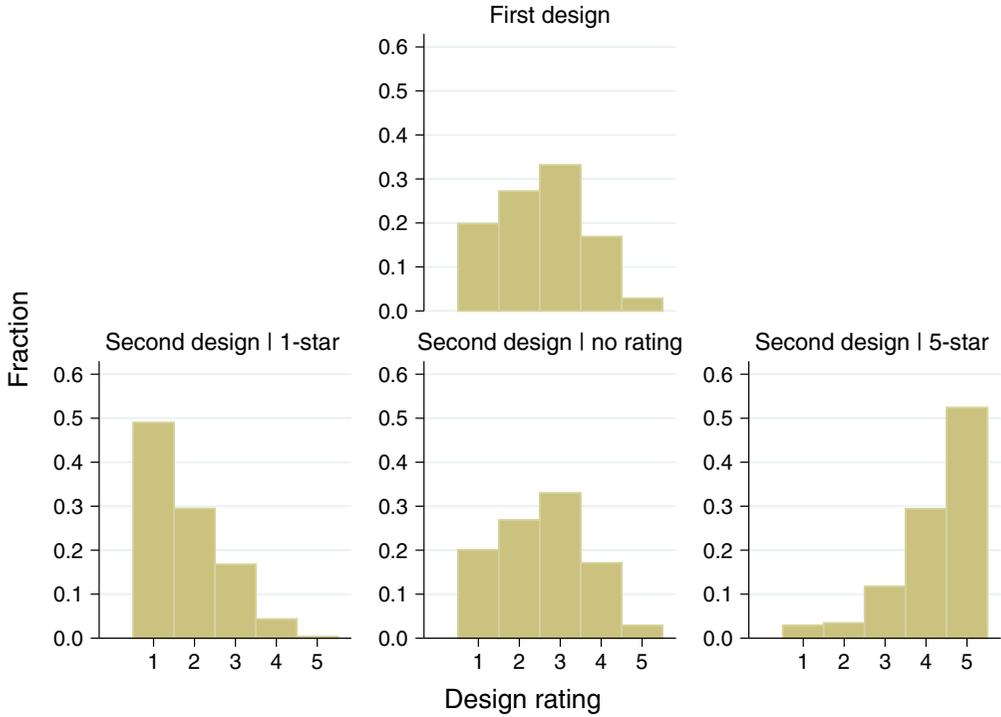
$$\tilde{s}_{ijk} = \left[ \left( \sum_{x < i} \mathbb{1}(\tilde{R}_{xjk} = \emptyset) > 0 \right), \left( \sum_{x < i} \mathbb{1}(\tilde{R}_{xjk} = 1) > 0 \right), \dots, \left( \sum_{x < i} \mathbb{1}(\tilde{R}_{xjk} = 5) > 0 \right) \right].$$

Figure 4 illustrates some examples of  $\hat{f}(\cdot)$ . The top panel shows the distribution on a player’s first design in a contest, and the bottom panel shows the distribution on a player’s second design, conditional on the first design receiving (from left to right): 1-star, no rating, and 5-stars. The results are intuitive: players with high ratings enter better designs, players with low ratings enter worse designs, and players with no feedback draw from the same distribution with their second design as with their first. I thus assume players know the distributions  $f(\cdot | \tilde{s})$  (or can infer them from intuition, experience, and examination of past contests) and plug in their observable ratings history,  $\tilde{s}_{ijk}$ , when they do the cost-benefit calculation to decide whether to continue participating.

Heterogeneity in ability likely exists even prior to feedback. To account for this possibility, I model players with no feedback as drawing from ability-specific distributions. For these cases, I estimate these distributions conditional on players’ quartile for average ratings in previous contests, adding an additional category for players who have no previous contests. In effect, this allows players with a good track record to draw their first design from a higher-quality bucket. However, once players have feedback on designs in a given contest, the estimation only

FIGURE 4

EXAMPLES OF NONPARAMETRIC DISTRIBUTIONS  $\hat{f}(\cdot)$  [Color figure can be viewed at wileyonlinelibrary.com]



Notes: Figure shows distribution of ratings for: players’ first design in a contest (top) and players’ second design after receiving a 1-star rating on their first design (bottom left), after no rating (bottom center), and after a 5-star rating (bottom right).

conditions on feedback received in that contest, and players’ track record in prior contests is no longer relevant to the estimation.

The expected rating on an additional design  $I_{jk} + 1$ , at the time player  $j$  makes it, is:

$$E[R_{I_{jk}+1,jk}] = \sum_r r \cdot f(r|\tilde{s}_{I_{jk}+1,jk}),$$

the weighted average of all possible ratings, weighted by the probability of each. The expected increase in the player’s odds of winning from the additional design, in a committed-prize contest and holding the competition constant, can similarly be written as follows:

$$\begin{aligned} E[\Delta Pr(j \text{ wins } k)] &= \sum_r (\Delta Pr(j \text{ wins } k)|R_{I_{jk}+1,jk} = r) \cdot f(r|\tilde{s}_{I_{jk}+1,jk}) \\ &= \sum_r \left( \frac{e^{\beta r} + \sum_{i \in I_{jk}} e^{\psi_{ijk}}}{e^{\beta r} + \sum_{i \in I_k} e^{\psi_{ik}}} - \frac{\sum_{i \in I_{jk}} e^{\psi_{ijk}}}{\sum_{i \in I_k} e^{\psi_{ik}}} \right) \cdot f(r|\tilde{s}_{I_{jk}+1,jk}). \end{aligned}$$

The first term in the parentheses is the probability of winning with an additional design rated  $r$ , whereas the second term is the probability of winning without it. Their difference is the increase in player  $j$ ’s odds from that design, which is weighted by the probability of an  $r$ -star

rating and summed to get its expected value. The expected benefit of the design is this incremental probability multiplied by the prize:

$$E[\text{MB}_{I_{jk+1},jk}] = E[\Delta \text{Pr}(j \text{ wins } k)] \cdot P_k.$$

*Estimating costs from stopping choices.* Having obtained a success function from the sponsor's choice problem, and derived a nonparametric procedure for predicting quality, estimation requires two final assumptions:

- Players exert effort if the expected benefit exceeds the cost.
- Players do not exert effort if the cost exceeds the benefit.

The appeal of these assumptions is self-evident: they impose minimal requirements on the agents and are nearly axiomatic in economic modelling. To make the logic concrete, consider the final moments of a contest. If designs can be made and entered in an infinitesimal amount of time (an assumption which, although hyperbolic, is perhaps not a bad approximation), then a contest should end in Nash equilibrium: given her ratings history and the ratings on competing designs, no player wants to enter any further designs. I similarly assume that a player's final design is a best response to competitors' play.

The implication is that the expected benefit of each player's final design exceeds her cost, whereas the expected benefit of an additional design does not. These conditions allow me to place bounds on costs. Given the level of participation in these contests, the incremental benefit of an additional entry is usually small, and the estimated bounds therefore tight. I thus assume each player's contest-specific effort cost to be bounded below by the expected benefit of the "extra" design that she does not to make ( $I_{jk} + 1$ ) and bounded above by the *ex ante* expected benefit of her final design ( $I_{jk}$ ), as follows:

$$C_{jk} \in [E[\text{MB}_{I_{jk+1},jk}], E[\text{MB}_{I_{jk},jk}]]$$

*Bootstrapped standard errors.* As functions of the MLE parameters and nonparametric frequencies, the estimated bounds are themselves random, taking the distribution of the convolution of their components. The maximum-likelihood estimates are known to be normally distributed. A player's predicted success function at a given vector of ratings is thus the ratio of a sum of log-normals over a sum of log-normals. This ratio is calculated for the "final" design, subtracted from a similar quantity at the "extra" design, multiplied by the nonparametric probability of a given rating, and summed over all possible ratings to obtain the bounds. Randomness will therefore enter from two sources: the MLE parameters and the nonparametric frequencies.

I use a block-bootstrap to obtain standard errors. To do so, I subsample entire contests from the data set with replacement, reestimate the logit parameters and nonparametric frequencies within the subsample, and use these estimates to recalculate bounds on cost for every contest-player in the original data set. My baseline bootstrap consists of 200 replications. As Section 5 shows, the bounds are estimated precisely, and the identified set for each contest-player's cost is narrow relative to the midpoint.

□ **Assumptions and identification.** Identification of players' design costs hinges on four assumptions:

- (i) The cost of effort is constant for a given player in a given contest.
- (ii) The latent quality of each design is linear in its rating and an i.i.d. logit error.
- (iii) The players know the distribution of ratings on their next design conditional on past ratings, as well as the process generating the odds of each design winning.
- (iv) Players exert effort up to the point where  $E[\text{MB}] \leq \text{MC}$ .

The first assumption is more an approximation than an assumption per se. The second assumption implies that all available information about quality is captured by a design's rating, and the reason the top-rated design does not win every contest boils down to luck: in practice, the

**TABLE 7** Summary Statistics for Estimated Costs

	<i>N</i>	Mean	Standard Deviation	Minimum	P25	P50	P75	Maximum
Est. Cost	160,059	3.86	3.28	0.04	1.44	3.06	5.34	78.09

Notes: Table provides summary statistics on the distribution of effort costs estimated by the model developed in Section 4. See text for further discussion.

sponsor may change its mind, or different people might administer ratings versus award winners. Although the third assumption can be debated, these distributions are both intuitive and available to any player that has competed in or browsed past contests. The fourth assumption derives from economic theory.

For the purposes of estimation, I further assume that:

- (v) At the time of their final submission, players have foresight over the state of the competition they will face at the end of the contest.

This final assumption is necessary for the agents to be able to compute the success function and evaluate the returns to effort. On average, the majority of players in a contest exit in the last quarter. Because the distribution of ratings in the contest is publicly available, I assume that players know or can forecast the competition they will face at its conclusion. Appendix D offers descriptive evidence that the information available midway through a contest is sufficient to project the state of competition at the end of the contest reasonably well, supporting an assumption of foresight.

## 5. Cost estimates

■ Table 7 provides summary statistics on the estimated costs of all 160,059 contest-players in the sample, which I calculate as the midpoint of the bounds estimated by the procedure described above. The cost estimates range from near zero to as high as \$78.09, with a mean of \$3.86 and median of \$3.06. To put these numbers in context, the average prize value per design in the sampled contests is \$3.34 (median \$2.86, max \$58.13).

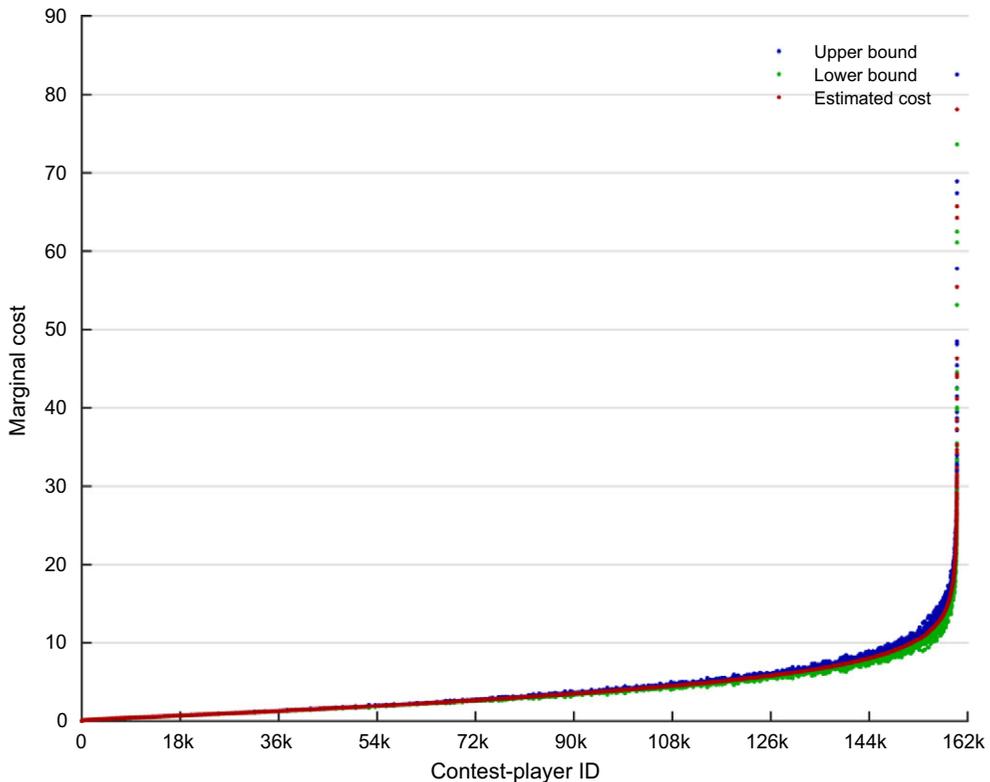
Figure 5 shows the estimated bounds on cost for every contest-player in the sample. The upper bound is shown in blue, and the lower bound in green. The red line traces the midpoint of these bounds for each player, which is my preferred point estimate for cost. Though confidence bands are not shown, these bounds are precisely estimated: the standard errors on the bounds are on average around 4.1% of the value (median 3.7%, 90th percentile 5.4%). They are also tight, with the width of the identified set on average being 2.3% of the midpoint (median 1.7%, 90th percentile 4.6%), further motivating the midpoint as a preferred point estimate.

Upon seeing these results, a natural question is whether the estimated costs are plausible. The mean cost is around the mean per-design winnings, suggesting that the estimates are sensible in magnitude. What of their variation? Closer inspection of contests where players are estimated to have very high costs typically reveals why. For example, in one such contest, players were asked to pay close attention to an unusually long list of requirements and provide a detailed, written explanation or story accompanying each design; as a result, only 23 designs were entered, in spite of a prize in the 99th percentile of all contests. This result is a direct consequence of the model: because the expected benefits to an additional design will be high when participation is low, *ceteris paribus*, costs must be high to rationalize the observed exit patterns. Examination of other contests yields similar intuition.

Although these estimates are inevitably approximations, their quality is evidenced by the fact that contest and player fixed effects explain nearly all (over 77%) of the variation in log costs, which should be the case if costs are primarily determined by the requirements of the contest and the characteristics of the player, and less so by the match between contests and players. Most

FIGURE 5

ESTIMATED DESIGN COSTS, WITH BOUNDS, IN ASCENDING ORDER [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]



Notes: Figure shows estimated costs for each contest-player in the sample. The upper bound is plotted in blue, and the lower bound in green. The red line traces the midpoint, which is the preferred point estimate. The figure arranges contest-players in increasing order of this midpoint.

of this variation (69%) is explained by contest fixed effects alone; in other words, costs vary considerably more between contests than within them. A smaller fraction (17%) is explained by player fixed effects alone.

To provide more systematic evidence, Table 8 correlates the mean, median, and standard deviation of the estimated costs in each contest with contest attributes described in previous sections. Costs tend to be higher with larger prizes, but lower when feedback is frequently provided. Intuitively, the results also suggest that sponsors will more readily commit to awarding large prizes in lower-cost (i.e., easier) projects.

Table 9 performs a similar exercise for contest-players, relating the estimated costs to past experience (prior contests entered) and performance (historical win percentage, average ratings). Each column moving left to right restricts to contest-players with successively greater experience, and all columns include contest and player fixed effects. The table shows that as players grow better (relative to other players and to their own past performance), they are estimated to have higher costs, consistent with higher-reservation wages and superior outside options. The table also shows a positive correlation between experience and costs, suggesting positive selection into extended participation on the platform—a pattern substantiated in nearly all longitudinal cuts of the data.

The foremost shortcoming of the cost estimates is the imprecision that results from the dimensionality reduction in the model. Recall that the empirical model has players drawing new

**TABLE 8** Correlation of Estimated Costs with Contest Attributes

	(1) Average Cost	(2) Median Cost	(3) Standard Deviation of Cost
Total prize value (\$100s)	0.541*** (0.085)	0.476*** (0.079)	0.271*** (0.029)
Committed value (\$100s)	-0.233** (0.105)	-0.203** (0.101)	-0.076** (0.039)
Fraction rated	-2.388*** (0.121)	-2.581*** (0.124)	-0.006 (0.041)
Contest length	-0.075*** (0.010)	-0.080*** (0.010)	-0.013*** (0.003)
Words in desc. (100s)	-0.020 (0.015)	-0.017 (0.014)	-0.014*** (0.005)
Attached materials	0.029 (0.022)	0.021 (0.023)	0.018** (0.008)
Prize committed	-0.111 (0.332)	-0.180 (0.320)	0.051 (0.120)
Constant	4.318*** (0.337)	4.057*** (0.319)	1.364*** (0.113)
<i>N</i>	4294	4294	4294
<i>R</i> <sup>2</sup>	0.18	0.18	0.13

Notes: Table shows the correlation of the features of the distribution of estimated costs with contest attributes. Observations are at the contest level. The dependent variable in Column (1) is the mean cost; Column (2), median cost; and Column (3), the standard deviation of costs. \*, \*\*, \*\*\* represent significance at the 0.1, 0.05, and 0.01 levels, respectively. Monthly fixed effects included but not shown. Robust standard errors in parentheses.

**TABLE 9** Correlation of Estimated Costs with Player Attributes

	(1) Ln(Cost)	(2) Ln(Cost)	(3) Ln(Cost)	(4) Ln(Cost)
Log(Prior contests)	0.026*** (0.008)	0.021*** (0.007)	0.021*** (0.008)	0.030*** (0.010)
Log(Historical win pct.)	-0.002 (0.010)	-0.009 (0.010)	-0.012 (0.011)	-0.009 (0.013)
Average rating	0.289*** (0.023)	0.366*** (0.026)	0.386*** (0.031)	0.423*** (0.044)
Constant	-0.253 (0.244)	-0.505** (0.247)	-0.674*** (0.248)	-0.823*** (0.264)
<i>N</i>	84132	81820	78406	67143
<i>R</i> <sup>2</sup>	0.75	0.75	0.75	0.76
Contest FEs	Yes	Yes	Yes	Yes
Player FEs	Yes	Yes	Yes	Yes
Experience	1	10	20	50

Notes: Table shows the correlation of the features of the estimated costs with player attributes. Observations are at the contest-player level. The dependent variable in all columns is log estimated cost. Each column restricts to contest-players with a minimum number of contests of prior experience, given in the final row of the table. \*, \*\*, \*\*\* represent significance at the 0.1, 0.05, and 0.01 levels, respectively. Standard errors clustered by player in parentheses.

designs and projecting outcomes on the basis of coarse ratings histories, with 64 variants. The low dimensionality is needed to make the estimation tractable, but it also sometimes results in multiple players having the same histories at the time of their stopping decisions and thus being estimated to have the same cost. This result is a reminder that the estimates are approximations, but it is not necessarily a reason to view them with skepticism: approximations will generally be sufficient to account for costs in regressions (Section 2) or to simulate counterfactuals (Section 6).

□ **Estimated costs are not strictly mechanical.** The nature of this procedure raises the question of whether the estimated costs are mechanical or substantive economic quantities. Recall that costs are estimated at the midpoint of the payoffs to a player's final design and an extra, unentered design ( $P_k \cdot \Delta\text{Pr}(\text{Win}|\text{final design})$  and  $P_k \cdot \Delta\text{Pr}(\text{Win}|\text{extra design})$ ). Holding fixed the  $\Delta\text{Pr}(\text{Win})$ , a 1% increase in prize would mechanically generate a 1% increase in the costs that I estimate. However, in practice, the increase in a player's probability of winning from entering an additional design will not be fixed: theory predicts that players will compete away the value of larger prizes. The probability gains available at the player's final or extra design will then be reduced, offsetting the mechanical effect of prize increases in the cost estimation.

To test for the presence of such an offset, Appendix Table E.1 regresses the log-probability gains achieved by a player's final design on the log prize. I find that larger prizes indeed tend to be heavily competed away: when the prize increases by 1%, the probability gains of players' final submissions decline by 0.75%—such that the residual cost estimate increases by only 0.25%. Though a perfect offset would manifest as an elasticity of -1, it should not be expected if projects with larger prizes are also more difficult. The previous results additionally show that costs relate to contest and player characteristics in expected ways, with sensible magnitude, reinforcing the evidence that the estimated costs are substantive economic quantities versus an artifact of the model.

□ **Total expenditure and overbidding.** The estimates also imply overbidding: total expenditures in a contest are often greater than the prize, consistent with an extensive body of experimental evidence from the tournament literature (Dechenaux, Kovenock, and Sheremeta, 2015). This overexpenditure may reflect any nonpecuniary value to participating (e.g., experience, portfolio-building, reputation) not fully captured in the cost estimates, and possibly errors in judgment. The dynamic setting might also provoke overinvestment via a sequence of marginal choices.

Appendix E examines implied total expenditures in detail. Figure E.1 shows the distribution of the ratio of total expenditures to prize value across contests, with a mean of roughly 140%. Table E.2 correlates this ratio with contest attributes and shows that overspending is especially common in contests with large and committed prizes, but declines sharply when players have more feedback (i.e., better information). Indeed, one doesn't need a structural model to see evidence of overbidding in high-value contests: as Figure E.2 shows, the number of contest submissions is strongly convex in the prize.

## 6. Counterfactual feedback mechanisms

■ I use these cost estimates to simulate alternative feedback policies. I first compare contest outcomes when (i) a random subset of submissions are rated, (ii) all submissions are rated, and (iii) ratings are private, relative to a baseline without feedback. I then activate selection and direction independently to compare their effects.

There are thus six simulated conditions: no feedback, partial feedback, universal feedback, private feedback, and selection enabled/improvement suppressed and vice versa.<sup>8</sup> Within each of these simulations, I sample 200 contests from the data and simulate (x50) three-player, sequential-play tournaments with an indefinite horizon. Limiting the field to only a few players reduces the dimensionality of the game sufficiently to be able to backward-induct the best responses of each player's competitors and allow players to internalize these best responses, imposing discipline on their choices. The disadvantage to this approach is that the simulated contests have substantially fewer players than those in the sample, limiting the role that selection can play on the extensive margin (number of players). However, because the intensive margin (number of designs) features prominently, selection will still be detectable, as the results will demonstrate.<sup>9</sup>

<sup>8</sup> Appendix F describes the implementation of each feedback condition in greater detail.

<sup>9</sup> The results are quantitatively similar when the simulations are expanded to four players (Appendix F). Note that all simulations sample the same contests to ensure comparability.

The formal procedure is as follows. For each simulation of a given contest, I first (randomly) select three players in the contest and fix the order of their moves.<sup>10</sup> The first player always enters. Beginning with player two, I project the distribution of ratings on her next design conditional on her prior ratings, as well as the distribution of ratings that player three and then player one might subsequently receive, should they choose to engage. I then (i) evaluate player one's decision to participate or abandon conditional on the preceding outcomes for players two and three, under a presumption of no further moves; (ii) backward-induct player three's decision to participate or abandon conditional on the preceding outcome for player two and the anticipated response of player one; and finally, (iii) backward-induct player two's decision to participate or abandon given the distribution of responses from her competitors, choosing the action with the higher-expected payoff.<sup>11</sup> If applicable, I then draw a design (in the form of a rating) for the given player, proceed to the next player, and repeat, iterating on this procedure until every player has exited.

When no feedback is provided, it is necessary to model players' beliefs over their relative performance. To do so, I insert an extra layer of simulation inside the continuation decision. I first seed potential performance histories after the same number of designs as observed at that point in the simulation. I then randomly draw histories for all three players, calculate the expected returns to continuation for the focal player, iterate 100 times, and take the average—thereby circumventing the absence of closed-form beliefs with computation. When feedback is private, I similarly insert an inner simulation of beliefs over competitors who (in this case) are expected to be improving over time.

As in the empirical model, feedback enters players' decision making in two places: it determines the distribution of the rating on a player's next design, and the value and shape of her success function. The direction effect surfaces in the former, whereas the selection effect is a convolution of the two: players remain active or abandon on the basis of the projected benefit to continuation, which is a function of both the rating on the next design and the incremental increase in the win probability that it generates. The net effects of feedback will reflect selection, direction, and their interaction, but simulation makes it possible to compare these effects by allowing one to vary while holding the other fixed, which is a possibility not otherwise supported by the data.

□ **Partial, full, and private feedback.** The first set of simulations compares outcomes when (i) a subset of designs are publicly rated, (ii) all are publicly rated, and (iii) all are privately rated, against the baseline of no feedback. In these simulations, each design will have a latent rating, and the variation between feedback policies is in the disclosure of this rating, which is implemented as follows. In the baseline, all ratings are concealed. With partial feedback, a random subset of designs have their ratings concealed, according to the empirical frequency of unrated designs in the sample, and these designs carry the same value of unrated designs as estimated in the data. With full feedback, all ratings are immediately revealed. With private feedback, ratings are revealed to the rated player but permanently concealed from competitors.

Table 10 shows the distribution (across contests) of total submissions under each policy, after taking the median across iterations of each contest as a representative value.<sup>12</sup> Total participation in these simulations is similar to the levels observed in the data—on average around 115 designs per contest—providing reassurance that the simulations are comparable, despite the reduction

<sup>10</sup> To further ensure comparability, the same players are selected in each iteration of each contest across mechanisms (i.e., every simulation of contest  $k$  always models the same players).

<sup>11</sup> This approach is taken due to the lack of a closed-form solution amenable to plug-in simulation. Foresight remains partially limited in this procedure, as the player two moves ahead is forecast to behave as if she is marginal, which is necessary to make the simulation dimensionality tractable.

<sup>12</sup> Throughout this section, I compare median (versus mean) outcomes for each contest to limit the influence of occasional outliers with extreme levels of participation on the results.

**TABLE 10** Distribution of Simulated Contests' Submission Counts under Each Feedback Policy

Simulated Policy	<i>N</i>	Mean	Standard Deviation	P10	P25	P50	P75	P90
No feedback	200	113.58	115.34	24.3	36.3	64.8	157.5	252.3
Partial feedback	200	112.50	120.50	21.8	33.0	63.8	147.5	259.3
Full feedback	200	107.25	116.86	20.5	30.8	60.3	134.0	249.5
Private feedback	200	120.58	123.28	23.8	37.0	69.0	165.3	281.5

Notes: Table shows the distribution of the median design count in each of 200 simulated contests, under each feedback policy: (i) no feedback; (ii) public feedback randomly provided to a subset of designs, according to the frequencies in the data; (iii) public feedback provided to all designs; and (iv) private feedback provided to all designs.

**TABLE 11** Effects of Feedback on Outcomes of Simulated Contests

Outcome	Percent Change in Outcome, When:		
	Some Rated (rel. to none rated)	All Rated (rel. to none rated)	Private Ratings (rel. to none rated)
Players	0.0% n.a. (0.0%)	0.0% n.a. (0.0%)	0.0% n.a. (0.0%)
Designs	-3.1%*** (0.8%)	-9.8%*** (0.7%)	4.9%*** (0.5%)
Num. 5-star	448.6%*** (22.4%)	551.0%*** (24.0%)	689.1%*** (22.8%)
Num. 4-star	66.3%*** (1.8%)	70.4%*** (1.9%)	98.5%*** (1.8%)
Num. 3-star	-11.3%*** (1.1%)	-23.6%*** (0.8%)	-8.0%*** (1.0%)
Num. 2-star	-43.7%*** (0.9%)	-59.7%*** (0.6%)	-47.3%*** (0.8%)
Num. 1-star	-64.0%*** (0.6%)	-75.7%*** (0.4%)	-67.3%*** (0.6%)

Notes: This table illustrates the effect of feedback on principal outcomes in simulated contests. A total of 200 contests were randomly selected from the sample, three players randomly chosen, and 50 three-player simulations performed for each contest. Simulations were performed under four scenarios: (i) no feedback; (ii) public feedback randomly provided to a subset of designs, according to the frequencies in the data; (iii) public feedback provided to all designs; and (iv) private feedback provided to all designs. In all cases, feedback is made available immediately after the player submits the design. In the private feedback scenario, players know the number of submissions from each competitor but not their ratings; when ratings are not known, the model simulates beliefs over relative performance. In order to reduce the effects of outlier simulations, I trim to median (rather than mean) simulated outcomes for each contest under each feedback policy. Columns show the average percent change in the given outcome relative to a baseline with no feedback. \*, \*\*, \*\*\* represent significance at the 0.1, 0.05, and 0.01 levels, respectively. Standard errors of these averages in parentheses.

to three players. Design counts decline as the frequency of feedback increases, but tick upward when ratings are private.

In Table 11, we see the effects of each policy on the number of players, designs, and designs of each rating. The table shows the average percent change in median outcomes versus the no-feedback baseline, with standard errors below. Given a pool of only three players, the number entering is unaffected, as they generally all participate. On the intensive margin, feedback reduces the number of submissions by 3% (partial) to 10% (full)—effects which are similar in magnitude to those in Table 3.

Drilling down, we see that feedback increases the number of top-rated designs and dramatically reduces the number of designs of every other rating: partial feedback on average increases the number of 5-star designs by a factor of 4.5, and full feedback by a factor of 5.5 (typically from a low base). Given these patterns, a natural next question is whether private feedback might do even better. The third column suggests this is indeed the case: despite that players expect their competitors to be improving, shrouding their precise performance generates a modest

**TABLE 12** Effects of Feedback on Outcomes of Simulated Contests

Outcome	Percent Change in Outcome, When:		
	Selection Only (rel. to no feedback)	Direction Only (rel. to no feedback)	Combined Effects (rel. to no feedback)
Players	0.0% n.a. (0.0%)	0.0% n.a. (0.0%)	0.0% n.a. (0.0%)
Designs	1.3%* (0.7%)	35.0%*** (0.7%)	-9.8%*** (0.7%)
Num. 5-star	14.9%*** (2.2%)	892.7%*** (29.3%)	551.0%*** (24.0%)
Num. 4-star	4.8%*** (1.4%)	137.8%*** (2.8%)	70.4%*** (1.9%)
Num. 3-star	0.1% (0.8%)	14.8%*** (1.2%)	-23.6%*** (0.8%)
Num. 2-star	-0.5% (0.8%)	-27.5%*** (1.3%)	-59.7%*** (0.6%)
Num. 1-star	-1.5%* (0.9%)	-46.7%*** (1.3%)	-75.7%*** (0.4%)

Notes: This table separates feedback's effects on quality and participation in the simulated contests. Effects are isolated by running simulations in which feedback is allowed to enter players' decisions to continue or drop out but not influence quality improvement (Column 1) and vice versa (Column 2). A total of 200 contests were randomly selected from the sample, three players randomly chosen, and 50 three-player simulations performed for each contest. In order to reduce the effects of outlier simulations, I trim to median (rather than mean) simulated outcomes for each contest under each feedback policy. Columns show the average percent change in the given outcome relative to a baseline with no feedback. \*, \*\*, \*\*\* represent significance at the 0.1, 0.05, and 0.01 levels, respectively. Standard errors of these averages in parentheses.

increase in 5-star submissions, with this growth fully attributable to an across-the-board increase in participation, as the relative frequency of each rating is unchanged.

□ **Decomposing selection and direction.** The second set of simulations toggles selection and direction to study their individual versus combined effects. The selection-only variant allows players to observe ratings on all designs at the continuation decision but does not allow them to improve. The direction-only variant does the reverse: players can leverage their own ratings to improve but only know the total number of designs in the contest, and not their ratings, when choosing whether to continue exerting effort. As in the other information-less conditions, players must then form beliefs over their own and their competitors' performance.

Table 12 provides the results. The first two columns show the selection and direction effects individually, and the third column shows their combined effect for comparison. When only selection is activated, the total number of designs is essentially unchanged, but the relative frequency of high-quality designs increases slightly (15%) by virtue of advantageous selection on modest *ex ante* differences in ability. When only direction is activated, the number of high-quality designs increases *ninefold* relative to no feedback. When both channels are activated, asymmetries drive down participation, producing a tempered—but still large—positive effect on the number of high-quality submissions and a sharp reduction in low-quality submissions. Together, the results imply that the effect of feedback on high-quality output is primarily attributable to improvement.

□ **Understanding the results.** The driving force behind differences in participation across feedback conditions is the development of asymmetry. In the simulations, as in the model, asymmetries reduce incentives for effort—either when a player is so strong that success is assured, or when competition is so severe that the player's returns to effort have been driven to zero. Feedback introduces asymmetries by making agents differentially productive, and by

clarifying the extent to which agents are ahead or behind and how productive each has become. Both pieces are necessary for asymmetry to have an effect, as players not only must develop heterogeneous performance and productivity, but these differences must be disclosed.

Asymmetries are why full feedback tends to yield fewer submissions than partial feedback, which in turn yields fewer submissions than no feedback: players diverge more rapidly, and this divergence is public knowledge. It also explains why the selection-only variant has essentially no effect on total submissions relative to no feedback: without improvement, players will remain about as symmetric or asymmetric as when the game began. Conversely, it can also explain the growth in submissions in the direction-only variant: agents do not know their precise performance when they must evaluate continuation at each round, and perceive themselves to grow similar over time, with any *ex ante* differences washed out by improvement. The intuition for why private feedback increases submissions relative to the no feedback condition is similar: players are improving and know that their competitors are as well, but precise performance differences are shrouded.

Incentives can be especially impacted by high ratings. When competitors' performance is public, these shocks create asymmetries that can bring the game to a swift conclusion—and the greater the asymmetry, the larger the impact. To demonstrate this, I examine changes in the returns to effort around the time that each player exits. I calculate the reduction in the returns to effort between a player's final submission and their next turn (at which they chose to abandon), normalize its value by the analogous change between their second-to-last and final submission, and take the median for each player across simulations. This quantity is a measure of the "shock": how much did the returns to effort typically decline immediately prior to abandonment, relative to the pace at which they had previously been declining? Values below one will indicate that returns declined gradually toward costs until the player dropped out, whereas values much greater than one indicate that abandonment was precipitated by a shock. Appendix Figure F.1 provides the distribution of these ratios across all 600 ( $= 200 \cdot 3$ ) simulated contests and players by feedback policy, confirming that when competition is shrouded, the ratio is around or below one, whereas when competitors' ratings are revealed, the values are greater and more dispersed.

## 7. Conclusion

■ This article studies the effects of performance feedback in competitive environments. Although feedback is widely thought to be an essential feature of product development, and is also widely used in organizational settings, its dual effects on productivity and incentives presents an intricate problem for the mechanism designer: feedback provides meaningful information that both generates quality or productivity gains and creates asymmetries between competitors that dampen incentives for effort. To my knowledge, this trade-off has not previously been identified or studied in the economics literature.

Using large-sample field data from commercial logo design competitions, I provide reduced-form evidence of this trade-off, develop a procedure to estimate structural parameters, and simulate counterfactual feedback mechanisms. The simulations suggest that feedback significantly increases high-quality output, with gains in quality far outweighing the costs to participation. Making feedback private yields modest incremental benefits by shrouding information on competitors' performance, which in turn reduces attrition. I then show that the positive effects of feedback on the quality of innovation in this setting are entirely the consequence of improvement, rather than selection on talent or luck. It is also likely that this latter finding is externally valid: in many creative activities, including research, learning-by-doing and perseverance are crucial to achieving success.

In light of the evidence in this article, a natural opportunity for future research is to further explore the optimal frequency of feedback and differences between private and public feedback mechanisms. An additional opportunity might be to compare a feedback mechanism to

elimination, which is widely used in practice to pare down the competition and sharpen incentives for remaining finalists. Finally, this article introduces a generalizable, structural approach to empirical research on tournaments, which has historically been constrained by a scarcity of large-sample data, and the methods developed in this article can be used or adapted to study tournaments in smaller samples.

## References

- AOYAGI, M. "Information Feedback in a Dynamic Tournament." *Games and Economic Behavior*, Vol. 70 (2010), pp. 242–260.
- AZMAT, G. AND IRIBERRI, N. "The Importance of Relative Performance Feedback Information: Evidence from a Natural Experiment Using High School Students." *Journal of Public Economics*, Vol. 94 (2010), pp. 435–452.
- BAIK, K.H. "Effort Levels in Contests with Two Asymmetric Players." *Southern Economic Journal*, Vol. 61 (1994), pp. 367–378.
- BEER, M. "Performance Appraisal." In J.W. Lorsch, ed., *The Handbook of Organizational Behavior*. Englewood Cliffs, NJ: Prentice Hall, 1987.
- BIMPIKIS, K., EHSANI, S., AND MOSTAGIR, M. "Designing Dynamic Contests." Working Paper, Stanford Graduate School of Business, 2016.
- BOUDREAU, K.J., LACETERA, N., AND LAKHANI, K.R. "Incentives and Problem Uncertainty in Innovation Contests: An Empirical Analysis." *Management Science*, Vol. 57 (2011), pp. 843–863.
- BOUDREAU, K.J., LAKHANI, K.R., AND MENIETTI, M. "Performance Responses to Competition Across Skill-Levels in Rank Order Tournaments: Field Evidence and Implications for Tournament Design." *RAND Journal of Economics*, Vol. 47 (2016), pp. 140–165.
- BROWN, J. "Quitters Never Win: The (Adverse) Incentive Effects of Competing with Superstars." *Journal of Political Economy*, Vol. 119 (2011), pp. 982–1013.
- CHE, Y.K. AND GALE, I. "Optimal Design of Research Contests." *American Economic Review*, Vol. 93 (2003), pp. 646–671.
- CHOI, J.P. "Dynamic R&D Competition under 'Hazard Rate' Uncertainty." *RAND Journal of Economics*, Vol. 22 (1991), pp. 596–610.
- DECHENAUX, E., KOVENOCK, D., AND SHEREMETA, R.M. "A Survey of Experimental Research on Contests, All-Pay Auctions and Tournaments." *Experimental Economics*, Vol. 18 (2015), pp. 609–669.
- EDERER, F. "Feedback and Motivation in Dynamic Tournaments." *Journal of Economics & Management Strategy*, Vol. 19 (2010), pp. 733–769.
- EDERER, F. AND FEHR, E. "Deception and Incentives: How Dishonesty Undermines Effort Provision." Discussion Paper no. 3200, IZA, 2007.
- ERIKSSON, T., POULSEN, A., AND VILLEVAL, M.C. "Feedback and Incentives: Experimental Evidence." *Labour Economics*, Vol. 16 (2009), pp. 679–688.
- FUDENBERG, D., GILBERT, R., STIGLITZ, J., AND TIROLE, J. "Preemption, Leapfrogging, and Competition in Patent Races." *European Economic Review*, Vol. 22 (1983), pp. 3–31.
- GERSHKOV, A. AND PERRY, M. "Tournaments with Midterm Reviews." *Games and Economic Behavior*, Vol. 66 (2009), pp. 162–190.
- GILL, D. "Strategic Disclosure of Intermediate Research Results." *Journal of Economics & Management Strategy*, Vol. 17 (2008), pp. 733–758.
- GOLTSMAN, M. AND MUKHERJEE, A. "Interim Performance Feedback in Multistage Tournaments: The Optimality of Partial Disclosure." *Journal of Labor Economics*, Vol. 29 (2011), pp. 229–265.
- GROSS, D.P. "Creativity under Fire: The Effects of Competition on Creative Production." Working Paper no. 16-109, Harvard Business School, 2016.
- HAILE, P.A. AND TAMER, E. "Inference with an Incomplete Model of English Auctions." *Journal of Political Economy*, Vol. 111 (2003), pp. 1–51.
- HALAC, M., KARTIK, N., AND LIU, Q. "Contests for Experimentation." *Journal of Political Economy*, (forthcoming).
- LAZEAR, E. AND ROSEN, S. "Rank-Order Tournaments as Optimum Labor Contracts." *Journal of Political Economy*, Vol. 89 (1981), pp. 841–864.
- MANSO, G. "Motivating Innovation." *The Journal of Finance*, Vol. 66 (2011), pp. 1823–1860.
- McFADDEN, D. "The Measurement of Urban Travel Demand." *Journal of Public Economics*, Vol. 3 (1974), pp. 303–328.
- MIHM, J. AND SCHLAPP, J. "Sourcing Innovation: On Feedback in Contests." Working Paper 2016/61/TOM, INSEAD, 2016.
- MURPHY, K.R. AND CLEVELAND, J.N. *Performance Appraisal: An Organizational Perspective*. Boston, MA: Allyn and Bacon, 1995.
- RIECK, T. "Information Disclosure in Innovation Contests." Discussion Paper no. 16, Bonn Graduate School of Economics, University of Bonn, 2010.
- TAYLOR, C.R. "Digging for Golden Carrots: An Analysis of Research Tournaments." *American Economic Review*, Vol. 85 (1995), pp. 872–890.

- TERWIESCH, C. AND XU, Y. "Innovation Contests, Open Innovation, and Multiagent Problem Solving." *Management Science*, Vol. 54 (2008), pp. 1529–1543.
- TULLOCK, G. "Efficient Rent Seeking." In J. Buchanan, R. Tollison, and G. Tullock, eds., *Toward a Theory of the Rent-Seeking Society*. College Station, TX: Texas A&M University Press, 1980.
- White House Office of Science and Technology Policy (OSTP). "Implementation of Federal Prize Authority: Fiscal Year 2013 Progress Report." 2014.
- WILLIAMS, H. "Innovation Inducement Prizes: Connecting Research to Policy." *Journal of Policy Analysis and Management*, Vol. 31 (2012), pp. 752–776.
- WIRTZ, J. "Feedback and Learning in Tournaments." Working Paper, Department of Economics, University of Warwick, 2014.