

# CREATIVITY UNDER FIRE: THE EFFECTS OF COMPETITION ON CREATIVE PRODUCTION

Daniel P. Gross\*

*Abstract*—Though fundamental to innovation and essential to many industries and occupations, individual creativity has received limited attention as an economic behavior and has historically proven difficult to study. This paper studies the incentive effects of competition on individuals' creative production. Using a sample of commercial logo design competitions and a novel, content-based measure of originality, I find that intensifying competition induces agents to produce original, untested ideas over tweaking their earlier work, but heavy competition drives them to stop investing altogether. The results yield lessons for the management of creative workers and the implementation of competitive procurement mechanisms for innovation.

## I. Introduction

THE creative act is a broadly important but understudied phenomenon in economics. Millions of people in the United States alone work in fields where creativity is essential to job performance, such as research, engineering, and professional services—industries that are the engines of innovation and growth in modern developed economies. CEO surveys also show that executives' top concerns consistently include the creativity of their employees and pursuit of innovation within the firm. Despite its importance, the creative act itself has received limited attention as an economic behavior and has historically proven difficult to study due to the challenge of measuring creativity and relating it to variation in incentives.

This paper studies the incentive effects of competition on individuals' creative output, exploiting a unique field setting where creative activity and competition can be measured and related: tournaments for the design of commercial logos and branding. Using image comparison tools to measure originality, I show that intensifying competition both creates and destroys incentives for creativity. While some competition is necessary to induce high-performing agents to develop original, untested designs over tweaking their existing work, heavy competition discourages effort of either kind. Theory suggests these patterns are driven by the risk-return trade-offs inherent in innovation. In the data, agents are most likely to

produce original designs in a horse race against one other competitor of similar quality.

It is useful to begin with a definition: creativity is the act of producing ideas that are novel and appropriate to the goal at hand (Amabile, 1996; Sternberg, 2008). The paper opens with a simple model that provides a framework for thinking about the economics of creative activity in a tournament setting, which both guides the empirical analysis and rationalizes its results.<sup>1</sup> In this model, a principal seeks a new product design and solicits candidates from a pool of workers via a tournament, awarding a prize to the best entry. Workers enter designs in turns, and once entered, each submission's quality is public knowledge. At each turn, workers must choose between developing an original design or tweaking a previous entry, cast as a choice between an uncertain and predictable outcome. The model suggests that competition increases workers' incentives to produce original designs over tweaks, but it also shows that heavy competition depresses incentives to do either. Though intuitive, and in part a recasting of prior theoretical research to this paper's context, the model is useful in framing and interpreting empirical results throughout the paper.

The paper then turns to an empirical study of logo design competitions, drawing on a sample of contests from a popular online platform.<sup>2</sup> In these contests, a firm ("sponsor") solicits custom designs from freelance designers ("players"), who compete for a winner-take-all prize. The contests in the sample offer prizes of a few hundred dollars and on average attract around 35 players and 100 designs. An important feature of this setting is that the sponsor can provide real-time feedback on players' designs in the form of one- to five-star ratings. These ratings allow players to gauge the quality of their own work and the intensity of competition while the contest is underway. Most important, the data set also includes the designs themselves, which makes it possible to study creative choices over the course of a contest. I use image comparison algorithms similar to those used by

Received for publication July 13, 2017. Revision accepted for publication January 7, 2019. Editor: Amit K. Khandelwal.

\*Gross: Harvard Business School and NBER.

I am grateful to Ben Handel, Steve Tadelis, and especially John Morgan, whose conversation provided much of the inspiration for this paper. I also thank colleagues at UC Berkeley and Harvard Business School, seminar and conference audiences, and the editor and anonymous reviewers for comments and suggestions that have improved the paper, as well as Branimir Dobeš for permission to include his work in this paper. Limited short segments of the text may be similar to passages from another of my papers ("Performance Feedback in Competitive Product Development"), which uses a different data set from the same setting. This research was supported by NSF grant DGE-1106400 as well as the Harvard Business School Division of Faculty and Research. All errors are my own.

A supplemental appendix is available online at [http://www.mitpressjournals.org/doi/suppl/10.1162/rest\\_a\\_00831](http://www.mitpressjournals.org/doi/suppl/10.1162/rest_a_00831).

<sup>1</sup>The model in this paper is related to Taylor (1995), Che and Gale (2003), Fullerton and McAfee (1999), and Terwiesch and Xu (2008) but differs in that it injects an explore-exploit dilemma into the agents' choice set. Whereas existing work models competing agents who must choose how much effort to exert, the agents in this paper must choose whether to build off an old idea or try a new one, much like a choice between incremental versus radical innovation. The framework also has ties to recent work on tournaments with feedback (Ederer, 2010), bandit problems in single-agent settings (Manso, 2011), and models of competing firms' choice over R&D project risk (Cabral, 2003; Anderson & Cabral, 2007).

<sup>2</sup>The empirical setting is conceptually similar to coding competitions studied by Boudreau, Lacetera, and Lakhani (2011), Boudreau, Lakhani, and Menietti (2016), and Boudreau and Lakhani (2015), though the opportunity to measure originality is unique. Wooten and Ulrich (2013, 2014) have also studied graphic design competitions, focusing on the effects of visibility and feedback.

commercial content-based image retrieval software (e.g., Google Image Search) to calculate similarity scores between pairs of images in a contest, which I then use to quantify the originality of each design relative to prior submissions by the same player and her competitors.

This setting presents a unique opportunity to observe creative production in the field. Though commercial advertising is important in its own right, the product development process observed here is similar to that in other domains where prototypes are created, tested, and refined. The nature of the setting enables a more detailed empirical study of this process, and its interaction with incentives, than is typically possible. The tournament format is especially germane. Although the website advertises itself as a crowd-sourcing platform, the contracting environment is fundamentally a request for proposals (RFP), a mechanism widely used by firms and government agencies to procure new products or technologies—often over multiple rounds, with interim scoring, and typically with only the top bid rewarded.

The sponsors' ratings are critical in this paper as a source of variation in the information that both the players and I have about the state of the competition. Using these ratings, I am able to directly estimate a player's probability of winning, and the results establish that ratings are meaningful: the highest-rated design in a contest may not always win, but a five-star design increases a player's win probability as much as 10 four-star designs, 100 three-star designs, and nearly 2,000 one-star designs. Data on the time at which designs are entered by players and rated by sponsors makes it possible to establish what every participant knows at each point in time. The empirical strategy exploits naturally occurring, quasi-random variation in the timing of sponsors' ratings and compares players' responses to information they observe at the time of design against that which is absent or not yet provided.

I find that competition has large effects on the content of players' submissions. Absent competition, positive feedback causes players to cut back sharply on originality: players with the top rating produce designs more than twice as similar to their previous entries as those with only low ratings. The effect is strongest when a player receives her first five-star rating—her next design will be a near replica of the highly rated design—and attenuates at each rung down the ratings ladder. However, these effects are reversed by half or more when high-quality competition is present: competitive pressure counteracts this positive feedback, inducing players to produce more original designs. A battery of supporting analysis establishes that this result is econometrically identified and is robust to alternative explanations and measures of the key variables.

Taken alone, these results suggest that competition unambiguously motivates creativity, but the analysis, and conclusion, presume no outside option. In practice, players have a third option: they can stop bidding. Whether and when this alternative becomes binding is its own question. Consistent with previous research from other tourna-

ment and tournament-like settings (Baik, 1994; Brown, 2011; Ross, 2012; Boudreau et al., 2016), I find that heavy competition discourages further investment. Empirically, high performers' tendency to produce original work is greatest when they face roughly fifty-fifty odds of winning—in other words, when neck-and-neck against one similar-quality competitor.

The driving assumption behind the model, and the interpretation of these results, is that creative effort is risky but high return. The data indicate that original designs outperform tweaks of low-rated work, but due to the ratings being bounded above at five stars, the same comparison cannot be made against tweaks of high-rated work. To test this assumption, I recruit a panel of professional designers to administer independent ratings designs on an extended scale and correlate their responses with these designs' originality. I find that original designs are on average more highly rated by these panelists than tweaks, but the distribution of opinion also has higher variance, reflecting risk. This evidence reinforces a possible link between creativity and risk taking suggested by research in other fields.

These findings contribute to a developing but mixed literature on the effects of competition on individual creative output: economists argue that competition can motivate the kind of risk taking that is characteristic of inventive activity (Cabral, 2003; Anderson & Cabral, 2007), yet many psychologists argue that high-powered incentives and other extrinsic pressures stifle creativity by crowding out intrinsic motivation (see Hennessey & Amabile, 2010, for a review) or by causing agents to choke (Ariely et al., 2009). Lab-based studies are as mixed as the theory (Eisenberger & Rhoades, 2001; Ederer & Manso, 2013; Erat & Gneezy, 2016; Charness & Grieco, 2018; Bradler, Neckermann, & Warnke, 2019), in part due to differences in measurement and experimental design. Missing from this literature is the added nuance that competition is not strictly a binary condition but rather can vary in intensity across treatments—and as this paper shows, the effects hinge crucially on the intensity of competition, as well as the existence of an outside option.

The evidence that creativity can be elicited with balanced competition has substantive implications for managers in creative industries and for the procurement practices of all organizations. Many practitioners appear to subscribe to the intrinsic motivation theory of creativity endorsed by social psychologists, which holds that extrinsic motivators are counterproductive, and regularly communicated in the *Harvard Business Review* (Florida & Goodnight, 2005; Amabile & Khaire, 2008; Amabile & Kramer, 2012) and other business publications. Although intrinsic motivation is valuable, the results of this paper show that high-powered incentives can be effective at motivating creativity. The results also provide lessons for organizers of innovation prize competitions and other competitive procurement mechanisms for innovation (e.g., RFPs) on managing the intensity of competition.

The paper also makes a methodological contribution to the innovation literature. Due to data constraints, empirical

research has historically measured innovation in terms of inputs (such as R&D spending) or outputs (patents), when innovation is at heart about the individual acts of discovery and invention that take place in between. As a result, there is relatively little systematic, empirical evidence on the process of idea production. This paper is an effort to fill this gap, invoking new tools for content-based measurement of innovation and using them to study how ideas are developed and refined in response to incentives.

The paper is organized as follows. Section I discusses related literature in economics and social psychology and presents the model. Section II introduces the empirical setting and describes the identification strategy. Section III estimates the effects of competition on the originality of submissions. Section IV presents the countervailing effects on participation. Section V provides evidence that creativity is risky but has a high return, supporting the key assumption of the model. Section VI discusses the implications of these results for policy, management, and future research on creativity and innovation and concludes.

## II. Background: Creativity and Incentives

### A. Existing Literature

Research on individual creativity has historically been a subject for social psychology. The question of whether incentives enhance or impair creativity is itself the focus of a contentious, decades-old debate led by two schools of thought: one camp argues that incentives impair creativity by crowding out intrinsic motivation (Amabile, 1996; Hennessey & Amabile, 2010), whereas the other argues that incentives bolster creativity, provided that creativity is explicitly what is being rewarded (Eisenberger & Cameron, 1996). Scholars in each of these camps have written public rejoinders to the other (Eisenberger & Cameron, 1998; Hennessey & Amabile, 1998), while others have sought to develop and test more nuanced theories in an attempt to reconcile these arguments (Shalley & Oldham, 1997).

The empirical literature on which these arguments are based in most cases invokes high-powered incentives (tournaments) in its experimental design. Despite dozens of experiments, the empirical evidence has been unable to clarify which of these positions is valid (Shalley, Zhou, & Oldham, 2004). Different papers include different-sized rewards (which may or may not be valuable enough to overcome motivational crowd-out, to the extent it occurs), different subject pools (college students versus grade-school children), and inconsistencies in how performance is evaluated and what features of performance are rewarded. Studies cited by the pro-incentives camp reward subjects for creativity, whereas studies cited by the anti-incentives camp evaluate creativity but often reward the best ideas. Experiments on both sides rely heavily on judges' assessments of creativity, which they are typically asked to score according to their own definitions.

Experimental economists have recently entered the literature, though often subject to the same limitations. Erat and Gneezy (2016) evaluate subjects' creativity in a puzzle-making task under piece-rate and competitive incentives and find that competition reduces creativity relative to an incentive-free baseline. Charness and Grieco (2019) in contrast find that high-powered incentives increase creativity in closed-ended creative tasks and have no effect on creativity in open-ended tasks. In both studies, creativity is scored by judges without guidance or a standardized definition, which leads to low interrater reliability. Rather than relying on subjective assessments, Bradler et al. (2019) study the effects of tournament incentives and gift exchange on creative output with an unusual uses task: subjects are asked to think of productive uses for a common household object (e.g., a tin can), and creativity is measured by the statistical infrequency of each answer. In this case, the authors find that tournaments increase creative output relative to both gift exchange and an incentive-free baseline, though the empirical methodology makes it hard to distinguish an increase in originality (novel uses) from an increase in output alone (total uses).

This paper makes several important departures from this body of research. The logo design competitions studied here provide a field setting in which actual creative professionals are competing for prizes of significantly greater value than observed in the existing lab-based studies. They also provide a setting where originality can be objectively measured with content-based assessment. Additionally, in contrast to much of the literature, it is not creativity per se that is being rewarded but product quality. As in most product development settings, creativity here is a means toward an end rather than an end in and of itself. Most important, however, this paper studies competition as a continuously varying rather than binary treatment. In practice, competition is not a uniform condition, and the fact that the implementation of competitive incentives varies across the previously cited studies might perhaps even explain their divergence.

At the heart of this paper is a set of empirical results on the originality of submissions into the sampled tournaments. To the extent that being creative is risky and its outcome uncertain, as the model below will propose, the paper is also connected to the economics literature on choices over risk in competition. Cabral (2003) and Anderson and Cabral (2007) show that in theory, laggards will take actions with higher-variance outcomes, consistent with the intuition of needing a big hit to catch up or, in the extreme, of having nothing to lose. Similar behavior has been observed among investment fund managers, who increase fund volatility after a midyear review that reveals them trailing their peers' performance (Brown, Harlow, & Starks, 1996) or when trailing the market (Chevalier & Ellison, 1997). In additional related work, Genakos and Pagliero (2012) study the choice over how much weight to attempt across rounds of dynamic weightlifting tournaments, and interpret the decision as a choice over risk. The authors find that whereas moderate laggards increase risk, distant laggards reduce risk—a result at odds with the existing

theoretical literature and the evidence from this paper, which indicate that more distant laggards prefer greater risk, conditional on participating. The interpretation, however, may be limited by the difficulty of empirically distinguishing a choice of risk from a commitment to a specified level of effort in the weightlifting context.

An additional literature to which this paper relates is the long-running literature in economics on product market competition and innovation (see Gilbert, 2006, and Cohen, 2010, for summaries). Since Schumpeter's (1942) contention that market power is favorable to innovation, researchers have produced explanations for and evidence of positive, negative, and inverted-U relationships between competition and innovation in a variety of markets—though the literature is complicated by differences in definition and measurement, challenges in econometric identification, and institutional variation. In a seminal contribution, Aghion et al. (2005) predict an inverted-U effect of product market competition on step-by-step innovation, and Aghion et al. (2014) find support for the predictions of this model in a lab experiment designed to mimic its features. There are, however, a few key differences between this paper's setting and the Aghion et al. (2005) model, the most important of which are the emphasis on individual creative behavior and the tournament context, where innovation is continuous and the intensity of competition is determined by relative performance differences rather than by an exogenous degree of collusion in the product market.

### B. Theoretical Framework

The preceding literature explains creativity and its motives primarily through narrative psychological constructs rather than economic forces. Yet creativity can also be interpreted as an economic behavior insofar as it involves a choice over uncertainty. This section demonstrates how this idea can be operationalized in a relatively simple tournament model whose features resemble the empirical setting. The results are presented in partial equilibrium to bring into focus the trade-offs facing agents in such a setting, which both guide the empirical analysis and offer a framework for interpreting the evidence that follows.

Suppose a risk-neutral principal seeks a product design. Because R&D outcomes are uncertain and difficult to value, the principal cannot contract directly on performance. It instead sponsors a tournament to solicit prototypes from  $J$  risk-neutral players, who enter designs sequentially and immediately learn of their quality. Each design can be either original or adapted from the blueprints of previous entries; players who choose to continue working on a given design at their next turn can reuse the blueprint to create variants, with the original version remaining in contention. At each turn, the player must decide whether to continue investing and, if so, whether to create an original design or tweak an earlier submission. At the end of the tournament, the sponsor awards a winner-take-all prize  $P$  to its favorite entry.

Let each design be characterized by latent value  $v_{jt}$ , which only the sponsor observes:

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

where  $j$  indexes players and  $t$  indexes designs. In this model,  $\beta_{jt}$  represents the design's quality, which is revealed by the sponsor's feedback, and the latent value is a function of revealed quality and an i.i.d. random shock, which reflects idiosyncrasies in the winner selection process. To hone intuition, further suppose that each player enters at most two designs. The type-1 extreme value error leads to logit choice probabilities for each design (see Train, 2009), such that player  $j$ 's total probability of winning is

$$\begin{aligned} Pr(\text{player } j \text{ wins}) &= \frac{\beta_{j0} + \beta_{j1}}{\beta_{j0} + \beta_{j1} + \sum_{k \neq j} (\beta_{k0} + \beta_{k1})} \\ &= \frac{\beta_{j0} + \beta_{j1}}{\beta_{j0} + \beta_{j1} + \mu_j}, \end{aligned} \quad (2)$$

where  $\mu_j \equiv \sum_{k \neq j} (\beta_{k0} + \beta_{k1})$  is the competition that player  $j$  faces in the contest. This function is concave in the player's own quality and decreasing in the quality of her competition.

Every player's first design in the contest is inherently novel, and entry is taken for granted. In theoretical terms, each player is endowed with her first submission. At their subsequent turn, players have three options: they can exploit (tweak, or adapt) the existing design, explore (experiment with) a radically different design, or abandon the contest altogether. To elaborate on each option:

- *Exploitation* costs  $c > 0$  and yields a design of the same quality, resulting in a second-round design with  $\beta_{j1} = \beta_{j0}$  and increasing the player's probability of winning accordingly.
- *Exploration* costs  $d \geq c$  and can yield a high- or low-quality design. With probability  $q$ , it will yield a high-quality design of  $\beta_{j1}^H = \alpha\beta_{j0}$ , and with probability  $(1 - q)$ , it will yield a low-quality design of  $\beta_{j1}^L = \frac{1}{\alpha}\beta_{j0}$ , where  $\alpha \geq 1$  is the exogenous degree of exploration.<sup>3</sup>
- *Abandonment* is costless. The player can abstain from further investment, leaving the player's probability of winning unchanged, as her previous work remains in contention.

In this context, feedback has three effects: it informs each player about her first design's quality, influences her second

<sup>3</sup>For the purposes of this illustrative model, I treat  $\alpha$  as fixed. If  $\alpha$  were endogenous and costless, the player's optimal  $\alpha$  would be infinite, since the exploration upside would then be unlimited and the downside bounded at 0. A natural extension would be to endogenize  $\alpha$  and allow exploration costs  $d(\cdot)$  or the probability of a successful outcome  $q(\cdot)$  to vary with it. Such a model is considerably more complex and beyond the scope of this paper.

design, and reveals the level of competition she faces. Players use this information to decide whether to continue participating and whether to do so by exploring a new design or reusing a previous one, which is a choice over which kind of effort to exert: creative or rote.

*Conditions for exploration.* To further simplify notation, let  $F(\beta_1) = F(\beta_1|\beta_0, \mu)$  denote a player's probability of winning when her second submission has quality  $\beta_1$ , given an initial submission of quality  $\beta_0$  and competition  $\mu$  (omitting the  $j$  subscript). For a player to produce an original design, she must prefer doing so over both exploiting the existing design and abandonment:

$$\underbrace{[qF(\beta_1^H) + (1-q)F(\beta_1^L)] \times P - d}_{E[\pi|_{\text{explore}}]} > \underbrace{F(\beta_0) \times P - c}_{E[\pi|_{\text{exploit}}]}, \quad (3)$$

$$\underbrace{[qF(\beta_1^H) + (1-q)F(\beta_1^L)] \times P - d}_{E[\pi|_{\text{explore}}]} > \underbrace{F(0) \times P}_{E[\pi|_{\text{abandon}}]}. \quad (4)$$

These conditions can be rearranged and written as follows:

$$qF(\beta_1^H) + (1-q)F(\beta_1^L) - F(\beta_0) > \frac{d-c}{P}, \quad (5)$$

$$qF(\beta_1^H) + (1-q)F(\beta_1^L) - F(0) > \frac{d}{P}. \quad (6)$$

In words, the probability gains from exploration over exploitation, and abandonment must exceed the difference in cost, normalized by the prize. If the difference in the cost of exploration versus exploitation is small relative to the prize, as it likely is in the data (see Gross, 2017), the choice between them reduces to a question of which choice yields the greater increase in the player's probability of winning.

*Effects of competition.* This modeling infrastructure leads directly to the focal propositions, which bring into focus how competition affects incentives for exploration.<sup>4</sup> To simplify the presentation, we will assume  $d = c$ , although the core result (that exploration is incentivized at intermediate levels of competition) also holds when  $d > c$ , with slightly more involved propositions, provided that  $d$  is not so high that exploration can never be optimal (see appendix A). The first proposition states that when  $\mu_j$  is high, exploration has greater expected benefits than exploitation, whereas when  $\mu_j$  is low, the reverse holds. The second proposition states that as  $\mu_j$  grows large, the benefits of a second design decline

<sup>4</sup>The propositions are provided in partial equilibrium (without strategic interactions) to emphasize the first-order trade-offs faced by agents in this setting. Strategic interactions, however, would not affect the result. At very small or large values of  $\mu$ , competitors' best responses will have little influence on the shape of the focal player's success function, and therefore little influence on the difference in returns to exploration versus exploitation. In the middle, there exists a threshold  $\mu^*$  that divides the real line into regions where exploration or exploitation yields greater benefits.

to 0. Because effort is costly, players are therefore likely to abandon the contest when competition grows severe.

**Proposition 1.** *Suppose  $q \in (\frac{1}{1+\alpha}, \frac{1}{2})$ . Then there exists a  $\mu^*$  such that for all  $\mu_j < \mu^*$ ,*

$$\underbrace{F(\beta_{j0})}_{E[\text{Pr}(\text{Win})|_{\text{exploit}}]} > \underbrace{[qF(\beta_{j1}^H) + (1-q)F(\beta_{j1}^L)]}_{E[\text{Pr}(\text{Win})|_{\text{explore}]},$$

and for all  $\mu_j > \mu^*$ ,

$$\underbrace{[qF(\beta_{j1}^H) + (1-q)F(\beta_{j1}^L)]}_{E[\text{Pr}(\text{Win})|_{\text{explore}]}} > \underbrace{F(\beta_{j0})}_{E[\text{Pr}(\text{Win})|_{\text{exploit}}]}.$$

**Proposition 2.** *The returns to a player's second design decline to 0 as  $\mu_j \rightarrow \infty$ .*

Proofs are provided in appendix A. The necessary condition for competition to motivate exploration is that  $q \in (\frac{1}{1+\alpha}, \frac{1}{2})$ , which holds if and only if original submissions are in expectation higher quality than tweaks, but successful outcomes are nevertheless improbable (see appendix A). In other words, exploration is not only risky but also high return. When this is the case, the first proposition shows that competition can provoke exploration as a strategic response, a result similar to the findings of Cabral (2003) and Anderson and Cabral (2007) on choices over risk, but in a structure more closely linked to the empirical setting: intuitively, when the player lags behind, the upside to exploration grows more valuable and the downside less costly. The second proposition shows, however, that large performance differences can also discourage effort, as the returns to effort decline to 0. The proposition is a reminder that participation must be incentivized. In contrast to many bandit models or models of choices over risk in competition (Cabral, 2003), agents in this setting incur costs and may withhold effort.<sup>5</sup>

Note that this setting may differ from other tournament settings in that submissions are cumulative: each player's first design remains in contention after the second is entered. The results of this section, however, would qualitatively remain the same if new designs replaced prior submissions: exploration would then be even riskier after a high-quality initial design and of similar risk after a low-quality initial design (as there is little to lose).<sup>6</sup> In either case, risk is defined by the ex ante variance of the success function and is the result of uncertainty over the quality of the second submission.

<sup>5</sup>Altogether, the model proposes that incentives for exploration are greatest at intermediate levels of competition (see appendix A). Mathematically, the result is driven by the curvature of the success function, which rises and then flattens with competition. Only at intermediate levels of competition does the function have adequate curvature to make the returns to exploration both larger than those to exploitation and large enough to exceed the cost.

<sup>6</sup>From a mechanism design perspective, this incremental risk could be mitigated if players were allowed to choose whether to replace the first submission with the new one after realizing the new submission's quality draw.

### III. Setting, Data, and Identification

I collect a randomly drawn sample of 122 logo design contests from a widely used online platform to study how creative behavior responds to competition.<sup>7</sup> The platform from which the data were collected hosts hundreds of contests each week in several categories of commercial graphic design, including logos, business cards, T-shirts, product packaging, book and magazine covers, and website and app mock-ups. Logos are the modal design category on this platform and thus a natural choice for analysis. A firm's choice of logo is also nontrivial, since it is the defining feature of its brand, which can be one of its most valuable assets and is how consumers will recognize and remember the firm for years to come.

In these contests, a firm (the sponsor—typically a small business or nonprofit organization) solicits custom designs from freelance designers (players) in exchange for a fixed prize awarded to its favorite entry. The sponsor publishes a project brief that describes its business, its customers, and what it likes and seeks to communicate with its logo; specifies the prize structure; sets a deadline for submissions; and opens the contest to competition. While the contest is active, players can enter (and withdraw) as many designs as they want, at any time they want, and sponsors can provide players with private, real-time feedback on their submissions in the form of one- to five-star ratings and written commentary. Players see a gallery of competing designs and the distribution of ratings on these designs, but not the ratings on specific competing designs. Copyright is enforced.<sup>8</sup> 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.

For each contest in the sample, I observe the project brief, which includes a project title and description, the sponsor's industry, and any specific elements that must be included in the logo; the contest's start and end dates; the prize amount; and whether the prize is committed (the sponsor may retain the option of not awarding the prize to any entries if none are to its liking). While multiple prizes are possible, the sample is restricted to contests with a single winner-take-all prize. I also observe every submitted design, the identity of the designer, his or her history on the platform, the time at 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 con-

attention, as sponsors can request that any withdrawn design be reinstated. Since I do not observe written feedback, I assume the content of written commentary is fully summarized by the rating.<sup>9</sup>

The player identifiers allow me to track players' activity over the course of each contest. I use the precise timing information to reconstruct the state of the 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 in terms of the feedback available (i.e., 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 I similarly require that feedback be provided at least one hour prior to the given design's submission to be considered observed at the time it is made.

The data set also includes the designs themselves. Recall that creativity bears the formal definition of the act of producing ideas that are novel and relevant to the goal at hand (Amabile, 1996; Sternberg, 2008). To operationalize this definition, I invoke image comparison algorithms commonly used in content-based image retrieval software (similar to Google Image's Search by Image feature) to measure the similarity of each design entered into a contest to preceding designs by the same and other players. I use two mathematically distinct procedures to compute similarity scores for image pairs, one of which is a preferred measure (the perceptual hash score) and the other of which is reserved for robustness checks (the difference hash score). Appendix B explains how they work. Each algorithm takes a pair of digital images as inputs; summarizes them in terms of a specific, structural feature; and returns a similarity score in the  $[0, 1]$  interval, with a value of 1 indicating a perfect match and 0 indicating total dissimilarity. This index effectively measures the absolute correlation of two images' underlying structure, reflecting similarities or differences in the basic shapes, outlines, and other elements that define the image.

To make this discussion concrete, figure 1 demonstrates an example application. The figure shows three designs, entered in the order shown, by the same player in a logo design competition that is similar to those in the sample, although not necessarily from the same platform.<sup>10</sup> The first two logos have some features in common (both use a circular frame and are presented against a similar backdrop), but they also have some stark differences. The perceptual hash algorithm gives them a similarity score of 0.31, and the difference hash algorithm scores them 0.51. The latter two logos are more alike, and though differences remain, they are now more limited and subtle. The perceptual hash algorithm gives these logos

<sup>7</sup>The sample consists of all logo design contests with public bidding that began the week of September 3–9, 2013, and every three weeks thereafter through the week of November 5–11, 2013, excluding those with multiple prizes or midcontest rule changes such as prize increases or deadline extensions. Appendix B describes the sampling procedures in greater detail.

<sup>8</sup>Though players can see competing designs, the site requires that all designs be original and enforces copyright protections. Players can also report violations if they believe a design has been copied or otherwise misused. Violators are permanently banned from the site. The site also prohibits the use of stock art and has a strict policy on the submission of overused design concepts. These mechanisms appear to be effective at limiting abuses.

<sup>9</sup>One of the threats to identification throughout the empirical analysis is that the estimated effects of ratings may be confounded by unobserved, written feedback: what seems to be a response to a rating could be a reaction to explicit direction provided by the sponsor. This concern is evaluated in detail in appendix D and discussed later in the paper.

<sup>10</sup>To keep the sampled platform anonymous, I omit identifying information.

FIGURE 1.—ILLUSTRATION OF IMAGE COMPARISON ALGORITHMS



The figure shows three logos entered in order by a single player in a single contest. The perceptual hash algorithm calculates a similarity score of 0.313 for logos 1 and 2 and a score of 0.711 for algorithms 2 and 3. The difference hash algorithm calculates similarity scores of 0.508 for algorithms 1 and 2 and 0.891 for algorithms 2 and 3.

a similarity score of 0.71, and the difference hash scores them 0.89.

For each design in a contest, I compute its maximal similarity to previous designs in the same contest by the same player. Subtracting this value from 1 yields an index of originality between 0 and 1, which can be interpreted as an empirical counterpart to the parameter  $1/\alpha$  in the model. In the empirical analysis, I primarily use measures of similarity to a player’s highest-rated previous submissions rather than all of her prior submissions, but since players tend to reuse only their highest-rated work, these two measures are highly correlated in practice ( $\rho = 0.9$  under either algorithm).

Creativity can manifest in this setting in other ways. For example, players sometimes create and enter several designs at once, and when doing so, they can make each one similar to or distinct from the others. To capture this phenomenon, I define “batches” of proximate designs entered into the same contest by a single player and compute the maximum intrabatch similarity as a measure of creativity in batched work. Two designs are proximate if they are entered within fifteen minutes of each other, and a batch is a set of designs in which every design in the set is proximate to another in the same set. Intrabatch similarity is an alternative measure of experimentation, reflecting players’ tendency to try minor variants of the same concept versus multiple concepts over a short period of time.

These measures are not without drawbacks or immune to debate. One drawback is that algorithmic comparisons require substantial dimensionality reduction and thus provide only a coarse comparison between designs based on a select set of features. Concerns on this front are mitigated by the fact that the empirical results throughout the paper are similar in sign, significance, and magnitude under two distinct algorithms. In addition, coarse comparisons will be sufficient for detecting designs that are plainly tweaks to earlier work versus those that are not, which is the margin that matters most for this paper. One may also question how well the algorithms emulate human perception, but the example provided above assuages this concern, as do other examples in appendix B, which discusses these issues in detail.

TABLE 1.—CHARACTERISTICS OF CONTESTS IN THE SAMPLE

Variable	N	Mean	SD	P25	P50	P75
Contest length (days)	122	8.52	3.20	7	7	11
Prize value (US\$)	122	247.57	84.92	200	200	225
Number of players	122	33.20	24.46	19	26	39
Number of designs	122	96.38	80.46	52	74	107
5-star designs	122	2.59	4.00	0	1	4
4-star designs	122	12.28	12.13	3	9	18
3-star designs	122	22.16	25.33	6	16	28
2-star designs	122	17.61	25.82	3	10	22
1-star designs	122	12.11	25.24	0	2	11
Unrated designs	122	29.62	31.43	7	19	40
Number rated	122	66.75	71.23	21	50	83
Fraction rated	122	0.64	0.30	0.4	0.7	0.9
Prize committed	122	0.56	0.50	0.0	1.0	1.0
Prize awarded	122	0.85	0.36	1.0	1.0	1.0

The table reports descriptive statistics for the contests. “Fraction rated” refers to the fraction of designs in each contest that is rated. “Prize committed” indicates whether the contest prize is committed to be paid (versus retractable). “Prize awarded” indicates whether the prize was awarded. The fraction of contests awarded subsumes the fraction committed, since committed prizes are always awarded.

TABLE 2.—DISTRIBUTION OF RATINGS (RATED DESIGNS ONLY)

	1-star	2-star	3-star	4-star	5-star	Total
Count	1,478	2,149	2,703	1,498	316	8,144
Percent	18.15	26.39	33.19	18.39	3.88	100

The table tabulates rated designs by rating: 69.3% of designs in the sample are rated by sponsors on a 1–5 scale. The site provides guidance on the meaning of each rating, which introduces consistency in the interpretation of ratings across contests.

A. Characteristics of the Sample

The average contest in the data lasts eight days, offers a \$250 prize, and attracts 96 designs from 33 players (table 1). On average, 64% of designs are rated, and less than three receive the top rating.

Among rated designs, the median and modal rating is three stars (table 2). Though fewer than 4% of rated designs receive a five-star rating, over 40% of all winning designs are rated five stars, suggesting that these ratings convey substantial information about a design’s quality and odds of success.<sup>11</sup>

<sup>11</sup>Another 33% of winning designs are rated four stars, and 24% are unrated.

TABLE 3.—SIMILARITY TO PRIOR DESIGNS BY SAME PLAYER AND COMPETITORS

Variable	N	Mean	SD	P10	P50	P90
A. Using Preferred algorithm: Perceptual hash						
Maximum similarity to any own prior designs	5,075	0.32	0.27	0.05	0.22	0.77
Maximum similarity to best own prior designs	3,871	0.28	0.27	0.03	0.17	0.72
Maximum similarity to best other prior designs	9,709	0.14	0.1	0.04	0.13	0.27
Maximum intrabatch similarity	1,987	0.45	0.32	0.05	0.41	0.91
B. Using alternative algorithm: Difference hash						
Maximum similarity to any own prior designs	5,075	0.58	0.28	0.16	0.62	0.94
Maximum similarity to best own prior designs	3,871	0.52	0.3	0.09	0.54	0.93
Maximum similarity to best other prior designs	9,709	0.33	0.21	0.09	0.29	0.63
Maximum intrabatch similarity	1,987	0.69	0.28	0.23	0.77	0.98

The table reports summary statistics on designs' similarity to previously entered designs (both own and competing). Pairwise similarity scores are calculated as described in the text and available for all designs whose digital image could be obtained (96% of entries; refer to the text for an explanation of missing images). The "best" preceding designs are those with the most positive feedback provided prior to the given design.

The website also provides formal guidance on the meaning of each star rating, which generates consistency in their interpretation and use across different sponsors and contests.

The designs themselves are available for 96% of submissions in the sample, and table 3 characterizes the similarity measures. For each design in the sample, we can compute its maximal similarity to previous designs by the same player, the highest-rated previous designs by the same player, and the highest-rated previous designs by that player's competitors (all in the same contest). For every design batch, I calculate the maximal similarity of any two designs in that batch.

The table shows that new entries are on average more similar to that player's own designs than her competitors' designs and that designs in the same batch tend to be more similar to each other than to previous designs by even the same player. But these averages mask more important patterns at the extremes. At the upper decile, designs can be very similar to previous work by the same player ( $\approx 0.75$  under the perceptual hash algorithm) or to other designs in the same batch (0.91), but even the designs most similar to competing work are not all that similar (0.27). At the lower end, designs can be original by all of these measures.

*Are ratings meaningful? Evidence from the empirical success function.* A simple cross-tabulation of ratings suggests that they are meaningful. In roughly two-thirds of contests, the winning design also had the highest rating in that contest (appendix table C3). But the relative value of each rating can be obtained by estimating an empirical success function, using the win-lose outcomes of each design in a large sample of contests from this platform, which I borrow from Gross (2017).<sup>12</sup> Recall from section II that a design's latent value is a function of its rating and an i.i.d. type I E.V. error. In the data, there are five possible ratings, such that this latent value can be flexibly specified with fixed effects for each rating and the success function estimated as a conditional logit. To formalize, let  $R_{ijk}$  denote the rating on design  $i$  by player  $j$

in contest  $k$ , and (in a slight abuse of notation) let  $R_{ijk} = \emptyset$  when design  $ijk$  is unrated. The value of each design,  $v_{ijk}$ , can be written as

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

The details of the estimation are provided in appendix C, but here we can summarize the results, which will be used in later analysis. The fixed effects are monotonically increasing in the rating and precisely estimated. Only a five-star design is on average preferred to the outside option. To produce the same increase in a player's estimated win probability as generated by a five-star design, a player would need 12 four-star designs, 137 three-star designs, or nearly 2,000 one-star designs. The magnitudes of these differences imply that for players with a top rating, competitive pressure primarily comes from other top-rated designs. As a measure of fit, the odds-on favorite wins almost half of all contests in the sample. With knowledge of the distribution of competitors' ratings, which is observable in practice, players can thus invoke a simple heuristic model similar to the one estimated here in their decision making.

### B. Empirical Methods and Identification

I exploit variation in the level and timing of ratings to estimate the effects of feedback and competition on players' creative choices. With time stamps on all activity, I can determine exactly what a player knows at each point in time about his or her own and about competitors' performance and identify the effects of performance differences known to players at the time of design. Identification is thus achieved by harnessing variation in the information players possess about the state of the competition.

Concretely, the analysis compares the actions of players who have received the top rating or know they face top-rated competition against those who do not—whether it is because no prior designs will be given the top rating or because these ratings have simply not yet been administered. The identifying assumption is that there are no omitted factors correlated

<sup>12</sup>Estimating the success function requires a larger sample of winners, and thus contests, than are in the primary sample of this paper. As appendix C shows, the sample in Gross (2017) contains more than 4,000 contests, is empirically comparable, and includes all of the same variables except for the images themselves—sufficient for the exercise.

with observed feedback that also affect choices. This assumption is supported by two pieces of evidence. First, the arrival of ratings is unpredictable, such that the set of ratings observed at any point in time is effectively random: sponsors are erratic, and it is difficult to know exactly when or how often a sponsor will log onto the site to rate new entries, much less whether or when they will rate any single design. More important, players' choices are uncorrelated with ratings that were unobserved at the time, including forthcoming ratings and ratings on specific competing submissions.<sup>13</sup>

To establish that feedback provision is unpredictable, I explore the relationship between feedback lags and ratings. In concept, sponsors may be quicker to rate the designs they like the most, to keep these players engaged and improving their work, in which case players might be able to infer their eventual ratings from the time elapsed without any feedback. Empirical assessment of this question (appendix D) confirms that this is not the case: whether measured in hours or as a percent of the total contest duration, the lag between when a design is entered and rated is unrelated to its rating. The probability that a rated design was rated before versus after the contest ends is similarly idiosyncratic.

Evidence that choices are uncorrelated with unobserved or not-yet-observed ratings is presented in section IV. Following the discussion of the focal results, I show that the relationship between the similarity measures and forthcoming ratings is indistinguishable from 0. In unreported tests, I also examine players' tendency to imitate highly rated competing designs and find no such patterns, likely because they simply do not know which designs are highly rated (and thus which ones to imitate). These collective results suggest that players respond only to ratings observed at the time of design.

Finally, a distinct threat to identification arises if ratings are accompanied by written feedback, and these comments provide explicit instruction that generates the patterns found in this paper. Appendix D evaluates this possibility in detail, using a newly drawn sample of contests in which written comments were made visible to the public, seemingly by error. Within this sample, sponsors provided written comments to fewer than 8% of submissions, though the frequency is significantly higher for highly rated submissions than for poorly rated ones. These comments take a range of flavors, with many echoing the rating given, but some make a more explicit request or suggest content changes. Because the latter present the risk of confounding the results of this paper, I hired individuals to read every comment in this sample and determine whether the sponsor suggested specific changes. Using this measure, I find that instructive comments are rare, not disproportionately provided to five-star designs relative to three- or four-star designs, and, most important, not related to the presence of high-rated competition, such that they can-

not be responsible for differences in behavior that will be found in the analysis below.

*Interpretation: Feedback versus market structure.* Although the phenomenological focus of the paper is on the effects of market structure on creativity, the identifying variation is generated by feedback, raising the question of whether the empirical results should be attributed to "feedback" or to "market structure"—though the distinction is blurred by the fact that in this setting, feedback *is* information about market structure. To orient this question, it is helpful to observe that feedback can affect choices through two channels in this setting. The rating on a given design will (a) help that player understand whether it is good or bad, and thereby project how her next design might be rated if she enters something similar to it versus different. But that rating, together with knowledge of the ratings given to competitors, will also (b) inform her whether her probability of winning is high or low—and if the player's objective is to win, this probability is fundamentally what will be driving behavior in these contests, as in the model in section II. In the following sections, I therefore refer to the effect of a player's own ratings as an effect of feedback or (for example) of a high rating, and to the effect of competitors' ratings as an effect of high-rated competition, with the primary focus being the latter.

#### IV. Competition and Creativity

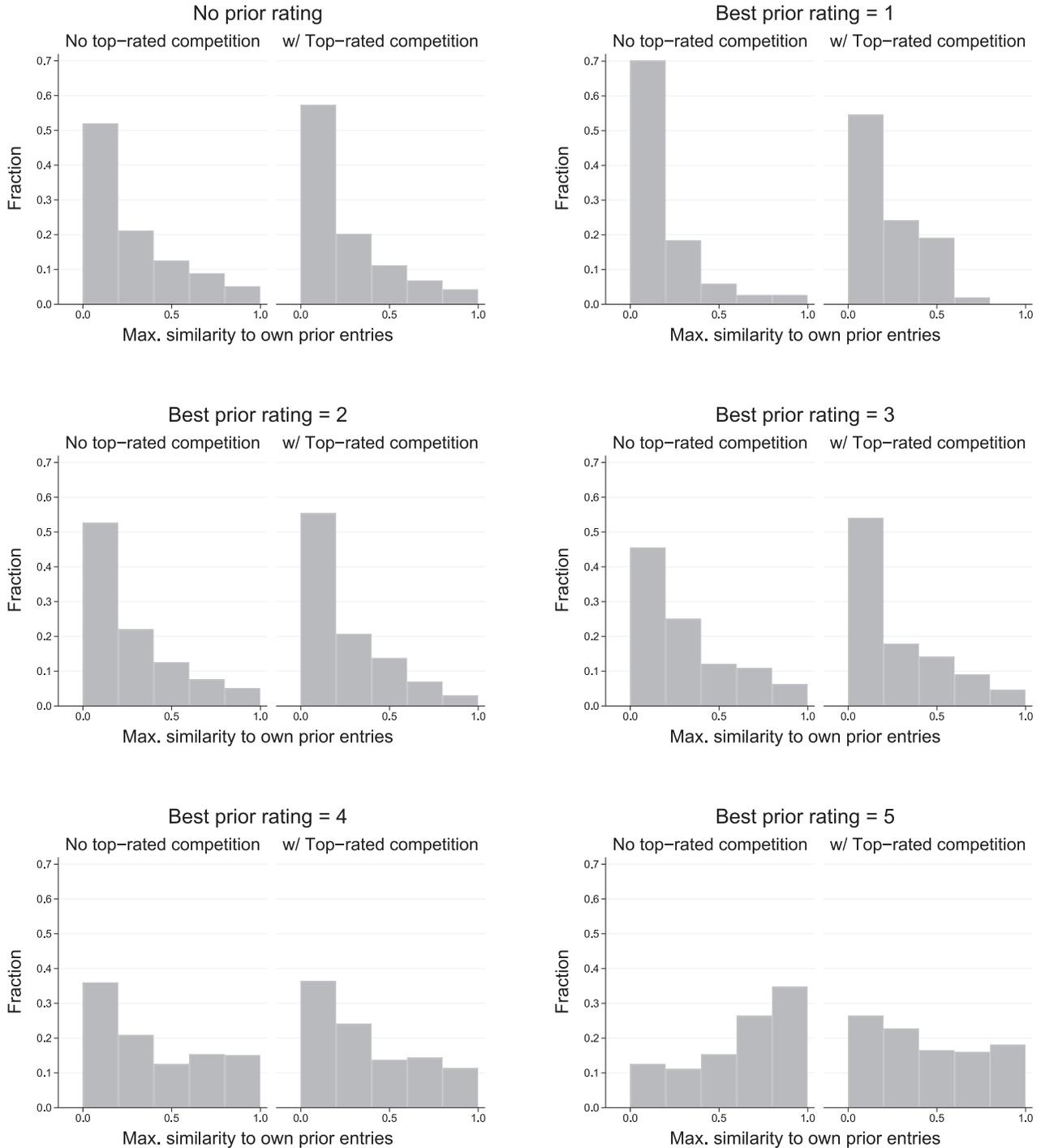
Figure 2 provides a first-cut, visual preview of this section's results. The figure shows the distribution of designs' maximal similarity to previous submissions by the same player, conditioning on that player's best rating and whether top-rated competition was present at the time. Recall that low similarity scores indicate that the submission is substantively original, while high scores indicate that it is a variant on a prior entry. Submissions from players with high ratings are significantly more similar to their prior work, with the effects largest for those with a five-star rating (bottom-right panel). However, these very same players are more likely to enter original designs when facing top-rated competition than in its absence.

The richness of the field setting makes it possible to evaluate whether competition affects high-performing players' tendency to enter more original work in a variety of ways. The estimating equation in this part of the paper is as follows, with variants estimated throughout this section and in the appendix:

$$\begin{aligned}
 \text{Similarity}_{ijk} = & \alpha + \sum_{r=2}^5 \beta_r \times \mathbb{1}(\bar{R}_{ijk} = r) \\
 & + \beta_{5c} \times \mathbb{1}(\bar{R}_{ijk} = 5) \times \mathbb{1}(\bar{R}_{-ijk} = 5) \\
 & + \beta_{5p} \times \mathbb{1}(\bar{R}_{ijk} = 5) \times P_k \\
 & + \beta_c \times \mathbb{1}(\bar{R}_{-ijk} = 5) + \beta_p \times P_k \\
 & + T_{ijk}\lambda + X_{ijk}\theta + \zeta_k + \varphi_j + \varepsilon_{ijk},
 \end{aligned}$$

<sup>13</sup>Though this setting may seem like a natural opportunity for a controlled experiment, the variation of interest is in the five-star ratings, which are sufficiently rare that a controlled intervention would require either unrealistic manipulation or an infeasibly large sample. I therefore exploit naturally occurring variation for this study.

FIGURE 2.—DISTRIBUTION OF NEW SUBMISSIONS' SIMILARITY TO OWN PRIOR SUBMISSIONS



The figure shows the distribution of designs' similarity to prior entries, conditional on the player's highest rating and the presence of top-rated competition. Each observation is a design, and the plotted variable is that design's maximal similarity to any previous submission by the same player in the same contest, taking values in [0,1], where a value of 1 indicates the design is identical to one of the player's earlier submissions.

where  $i$  indexes designs by player  $j$  in contest  $k$ .  $Similarity_{ijk}$  represents the maximal similarity of design  $ijk$  to the highest-rated preceding designs by player  $j$  in contest  $k$ ;  $\bar{R}_{ijk}$  is the highest rating received by player  $j$  in contest  $k$  prior to design  $ijk$ ;  $\bar{R}_{-ijk}$  is the highest rating received by player  $j$ 's competitors prior to design  $ijk$ ;  $P_k$  is the prize in contest  $k$

(in \$100s);  $T_{ijk}$  is the fraction of the contest elapsed when design  $ijk$  is entered;  $X_{ijk}$  consists of other design-level controls, including counts of previous designs by the same player and by competitors, and the number of days remaining in the contest; and  $\zeta_k$  and  $\varphi_j$  are contest and player fixed effects. Specifications without contest fixed effects include  $P_k$  as a

TABLE 4.—SIMILARITY TO PLAYER’S BEST PREVIOUSLY RATED DESIGNS

	(1)	(2)	(3)	(4)	(5)
Player’s prior best rating==5	0.440*** (0.102)	0.459*** (0.092)	0.260*** (0.097)	0.357*** (0.097)	0.362*** (0.102)
× 1+ competing 5-stars	-0.197*** (0.073)	-0.245*** (0.063)	-0.158** (0.061)	-0.206*** (0.070)	-0.208*** (0.071)
× prize value (\$100s)	-0.025 (0.025)	-0.015 (0.023)	0.005 (0.032)	-0.014 (0.031)	-0.018 (0.033)
Player’s prior best rating==4	0.165*** (0.024)	0.160*** (0.022)	0.128*** (0.030)	0.121*** (0.031)	0.116*** (0.032)
Player’s prior best rating==3	0.079*** (0.018)	0.077*** (0.018)	0.068** (0.028)	0.060** (0.028)	0.056** (0.028)
Player’s prior best rating==2	0.044** (0.021)	0.044** (0.022)	0.023 (0.029)	0.026 (0.030)	0.024 (0.030)
One or more competing 5-stars	-0.020 (0.018)	-0.009 (0.020)	-0.003 (0.022)	0.004 (0.023)	0.001 (0.024)
Prize value (\$100s)	-0.014* (0.007)		-0.010 (0.010)		
% of contest elapsed	-0.030 (0.034)	-0.060* (0.032)	-0.010 (0.030)	-0.018 (0.034)	-0.103 (0.084)
Constant	0.238*** (0.039)	0.207*** (0.023)	0.235*** (0.044)	0.232*** (0.061)	0.303*** (0.093)
N	3,871	3,871	3,871	3,871	3,871
R <sup>2</sup>	0.07	0.20	0.48	0.53	0.53
Contest FEs	No	Yes	No	Yes	Yes
Player FEs	No	No	Yes	Yes	Yes
Other controls	No	No	No	No	Yes

Observations are designs. The dependent variable is a continuous measure of a design’s similarity to the highest-rated preceding entry by the same player, taking values in [0,1], where a value of 1 indicates the design is identical to another. Column 5 controls for the number of days remaining and the number of previous designs entered by the player and her competitors. Similarity scores in this table are calculated using a perceptual hash algorithm. Preceding designs/ratings are defined to be those entered/provided at least sixty minutes prior to the given design. Significant at \*0.1, \*\*0.05, \*\*\*0.01. Standard errors clustered by player in parentheses.

stand-alone explanatory variable to identify the interaction. Standard errors throughout are clustered by player to account for any within-player correlation in the error term, though the results are robust to (and more conservative than) clustering by contest player or by contest.

A. *Similarity of New Designs to a Player’s Previous Designs*

I begin by examining players’ tendency to enter novel versus derivative designs. Table 4 provides estimates from regressions of the maximal similarity of each design to the highest-rated preceding designs by the same player on indicators for the highest rating received. Column 1 presents a baseline with no fixed effects or other controls. Columns 2 and 3 add fixed effects for contests and players, respectively, and column 4 includes both. Column 5 additionally controls for the number of days remaining in the contest and the number of prior submissions by the same player, as well as competing players.

Similar patterns are observed across all specifications. In the regression with both fixed effects and controls (column 5), we see that players with the top rating enter designs that are 0.36 points, or over 1 full standard deviation, more similar to their previous work than players who have only low ratings or no ratings. Roughly half of this effect is reversed by the presence of top-rated competition, with this counteracting effect significant at the 1% level. When a player’s highest rating is four stars, her new designs are on average around 0.1 point more similar to previous work. This effect further

attenuates as the best observed rating declines until it is indistinguishable from 0 at a best rating of two stars, with all such differences statistically significant. High-rated competition is not observed to have an effect on similarity for these lower performers, who are already unlikely to reuse their low-rated submissions.

The latter regressions in table 4 use contest and player fixed effects to control for other factors that are either common to all players within a contest or across all contests for a given player, but they do not control for factors that are constant for a given player within specific contests, as doing so leaves too little variation to identify the focal effects. Such factors could nevertheless be confounding, such as if players who continue participating in different competitive conditions are systematically more or less likely to enter similar designs in that contest. The estimates in the previous tables additionally mask potential heterogeneity in players’ reactions to competitive conditions over the course of a contest.

Table 5 addresses these concerns with a model in first differences. The dependent variable here is the change in designs’ similarity to the player’s best previously rated work. This variable can take values in [-1, 1], where a value of 0 indicates that the given design is as similar to the player’s best preceding design as was the last one she entered, and a value of 1 indicates that the player transitioned fully from innovating to recycling and a value of -1, the converse. The independent variables are changes in indicators for the highest rating the player has received, interacting the indicator for the top rating with the prize and the presence of top-rated

TABLE 5.—CHANGE IN SIMILARITY TO PLAYER'S BEST PREVIOUSLY RATED DESIGNS

	(1)	(2)	(3)	(4)	(5)
$\Delta(\text{Player's best rating}=5)$	0.861*** (0.162)	0.878*** (0.170)	0.928*** (0.203)	0.914*** (0.205)	0.924*** (0.205)
× 1+ competing 5-stars	-0.417*** (0.118)	-0.412*** (0.125)	-0.418*** (0.144)	-0.427*** (0.152)	-0.429*** (0.152)
× prize value (\$100s)	-0.092** (0.039)	-0.094** (0.039)	-0.115** (0.049)	-0.107** (0.047)	-0.110** (0.048)
$\Delta(\text{Player's best rating}=4)$	0.275*** (0.062)	0.282*** (0.065)	0.267*** (0.073)	0.276*** (0.079)	0.279*** (0.079)
$\Delta(\text{Player's best rating}=3)$	0.143*** (0.055)	0.151*** (0.058)	0.134** (0.065)	0.137** (0.069)	0.138** (0.069)
$\Delta(\text{Player's best rating}=2)$	0.079* (0.043)	0.082* (0.046)	0.063 (0.053)	0.059 (0.056)	0.059 (0.057)
One or more competing 5-stars	-0.003 (0.007)	-0.003 (0.015)	-0.003 (0.014)	0.004 (0.025)	0.003 (0.026)
Prize value (\$100s)	0.003 (0.002)		0.003 (0.008)		
% of contest elapsed	0.015 (0.012)	0.009 (0.018)	0.017 (0.024)	0.004 (0.030)	-0.048 (0.074)
Constant	-0.029*** (0.010)	-0.017* (0.010)	-0.031 (0.029)	0.063 (0.093)	0.105 (0.108)
<i>N</i>	2,694	2,694	2,694	2,694	2,694
<i>R</i> <sup>2</sup>	0.03	0.05	0.11	0.14	0.14
Contest FEs	No	Yes	No	Yes	Yes
Player FEs	No	No	Yes	Yes	Yes
Other controls	No	No	No	No	Yes

Observations are designs. The dependent variable is a continuous measure of the change in designs' similarity to the highest-rated preceding entry by the same player, taking values in  $[-1,1]$ , where a value of 0 indicates that the player's current design is as similar to her best preceding design as was her previous design, and a value of 1 indicates that the player transitioned fully from innovating to recycling (and a value of  $-1$ , the converse). Column 5 controls for the number of days remaining and number of previous designs entered by the player and her competitors. Similarity scores in this table are calculated using a perceptual hash algorithm. Preceding designs/ratings are defined to be those entered/provided at least sixty minutes prior to the given design. Significant at \*0.1, \*\*0.05, \*\*\*0.01. Standard errors clustered by player in parentheses.

competition. I estimate this model with the same configurations of contest fixed effects, player fixed effects, and controls to account for other potential reasons why players' propensity for similarity changes over time.

The results provide even stronger evidence of how competition affects creative choices and are statistically and quantitatively similar across specifications. A player who receives her first five-star rating will typically then enter a near replica: the similarity increases by 0.9 point, or over 3 standard deviations, relative to players with low ratings. Top-rated competition again reverses roughly half of this effect, with the difference significant at the 1% level. Given their magnitudes, these effects will be plainly visible to the naked eye (figure 1 in section III gives an example of what they would look like in practice). The effects of a new best rating of four, three, or two stars attenuate monotonically, similar to earlier results, and high-rated competition is not seen to have an effect on low performers.<sup>14</sup>

The appendix provides robustness checks and supplementary analysis. In appendix E, I show that similar patterns exist for experimentation within submission batches, and that they

also arise for four-on-four competition (players with four-star ratings and four-star competition) when no five-star ratings have been granted. Appendix E also examines the originality of players' initial submissions in a contest (measured against previous submissions by others in that contest, the same player in other contests, or everyone in all prior contests), which are mechanically excluded by the focal measure (which compares designs against prior submissions by the same player in the same contest) and finds no correlation with competition at the time of the initial submission, reinforcing that the effects documented in the tables above are specific to a player's portfolio of work within the given contest.

To confirm that these patterns are not an artifact of the perceptual hash algorithm, appendix F reestimates the regressions in the preceding tables using the difference hash algorithm to calculate similarity scores. The results are both statistically and quantitatively similar. In appendix G, I split out the effects of competition by the number of top-rated competing designs, finding no statistical differences between the effects of one versus more than one: all of the effects of competition are achieved by one high-quality competitor.

This latter result is especially important for ruling out an information-based story. In particular, the presence of other five-star ratings might indicate that the sponsor has diverse preferences and that unique designs have a higher likelihood of being well received than one might otherwise believe. If this were the case, then similarity should continue to decline

<sup>14</sup>Interestingly, these regressions also find that new recipients of a top rating can be induced to explore new designs by larger prizes. The theory suggests a possible explanation: large prizes moderate the influence of costs in players' decision making. If original designs are more costly (take more time or effort) than tweaks, they may be more worthwhile when the prize is large. This is particularly the case for players with highly rated work in the contest.

TABLE 6.—SIMILARITY TO PLAYER’S BEST NOT-YET-RATED DESIGNS (PLACEBO TEST)

	Similarity to forthcoming			Residual
	(1)	(2)	(3)	(4)
Player’s best forthcoming rating==5	0.007 (0.169)	-0.084 (0.136)	-0.105 (0.151)	-0.113 (0.122)
× 1+ competing 5-stars	-0.094 (0.099)	0.032 (0.056)	0.027 (0.066)	0.035 (0.062)
× prize value (\$100s)	-0.003 (0.031)	0.015 (0.025)	0.021 (0.027)	0.018 (0.025)
Player’s best forthcoming rating==4	0.039 (0.066)	0.051 (0.096)	0.049 (0.094)	0.034 (0.095)
Player’s best forthcoming rating==3	0.080 (0.052)	0.049 (0.088)	0.051 (0.088)	0.036 (0.088)
Player’s best forthcoming rating==2	0.030 (0.049)	-0.010 (0.093)	-0.007 (0.094)	-0.014 (0.095)
One or more competing 5-stars	-0.080 (0.097)	-0.013 (0.110)	-0.010 (0.117)	-0.013 (0.119)
% of contest elapsed	0.016 (0.242)	-0.502 (0.478)	-0.466 (0.462)	-0.468 (0.497)
Constant	0.217 (0.212)	0.556 (0.560)	0.569 (0.543)	0.398 (0.581)
N	1,147	577	577	577
R <sup>2</sup>	0.68	0.83	0.83	0.67
Contest FEs	Yes	Yes	Yes	Yes
Player FEs	Yes	Yes	Yes	Yes
Other controls	Yes	Yes	Yes	Yes

The table provides a placebo test of the effects of future feedback on similarity. Observations are designs. The dependent variable in columns 1 to 3 is a continuous measure of a design’s similarity to the best design that the player has previously entered that has yet to but will eventually be rated, taking values in [0,1] where a value of 1 indicates that the two designs are identical. Under the identifying assumption that future feedback is unpredictable, current choices should be unrelated to forthcoming ratings. Note that a given design’s similarity to an earlier, unrated design can be incidental if they are both tweaks on a rated third design. To account for this possibility, column 2 controls for the given and unrated designs’ similarity to the best previously rated design. Column 3 allows these controls to vary with the highest rating previously received. The dependent variable in column 4 is the residual from a regression of the dependent variable in the previous columns on these controls. These residuals will be the subset of a given design’s similarity to the unrated design that is not explained by a jointly occurring similarity to a third design. All columns control for days remaining and number of previous designs by the player and her competitors. Similarity scores in this table are calculated using a perceptual hash algorithm. Preceding designs/ratings are defined to be those entered/provided at least sixty minutes prior to the given design. Significant at \*0.1, \*\*0.05, \*\*\*0.01. Standard errors clustered by player in parentheses.

as five-star competitors are revealed. That this is not the case suggests that the effect is in fact the result of variation in incentives from competition.

In unreported tests, I also look for effects of five-star competition on players with only four-star designs and find attenuated effects that are negative but not significantly different from 0. I also explore the effects of prize commitment, since the sponsor’s outside option of not awarding the prize is itself a competing alternative. The effect of prize commitment is not statistically different from 0. I similarly test for effects of four-star competition on players with five-star designs, finding none. These results reinforce the earlier evidence that competition effectively comes from other designs with the top rating.<sup>15</sup>

*B. Placebo Test: Similarity to a Player’s Not-Yet-Rated Designs*

The identifying assumptions require that players are not acting on information that correlates with feedback but is unobserved in the data. As a simple validation exercise, the regressions in table 6 perform a placebo test of whether simi-

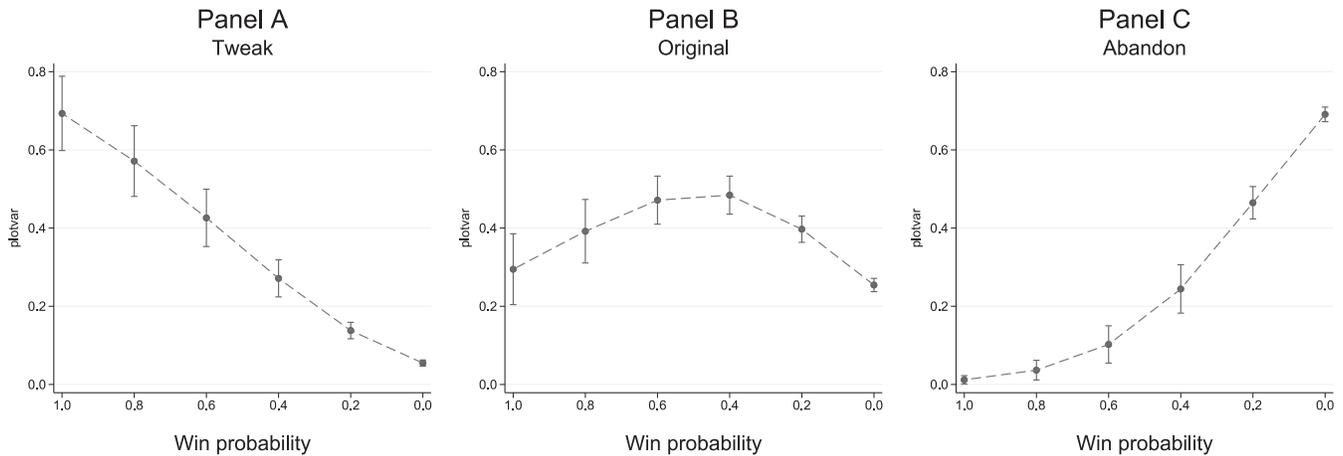
larity is related to impending feedback. If an omitted determinant of creative choices is correlated with ratings and biasing the results, then it would appear as if similarity responds to forthcoming ratings. If the identifying assumptions hold, we should see only 0s.

The specification in column 1 regresses a design’s maximal similarity to the player’s best designs that have not yet been but will eventually be rated on indicators for the ratings they later receive. I find no evidence that the similarity of designs is related to forthcoming ratings. Because a given design’s similarity to a not-yet-rated design can be incidental if both are tweaks on a third design, column 2 controls for similarity to the best already-rated design. Column 3 allows this control to vary with this existing rating. As a final check, I isolate the similarity to the not-yet-rated design that cannot be explained by similarity to the third design in the form of a residual, and in column 4, I regress these residuals on the same independent variables. In all cases, I find no evidence that players tweak designs with higher forthcoming ratings. Choices are only correlated with ratings observed in advance.

**V. Effects on Participation**

The analysis thus far conditions on continued participation: the unit of observation is a submission, and the outcome is its similarity to previous submissions. However, players can also stop making submissions if they perceive the returns to

<sup>15</sup>In additional unreported results, I also reestimate table 4 for players who entered the contest when the highest competing rating was four stars or higher versus three stars or lower, to see whether selection into the contest on the intensity of competition might explain the results. I find similar effects for both groups.

FIGURE 3.—PROBABILITY OF TWEAKS, ORIGINAL DESIGNS, AND ABANDONMENT AS A FUNCTION OF  $\text{Pr}(\text{WIN})$ 

The figure plots the probability that a player does one of the following on (and after) a given submission, as a function of his or her contemporaneous win probability: tweaks and then enters more designs (panel A), experiments and then enters more designs (panel B), or stops investing in the contest (panel C). These probabilities are estimated as described in the text, and the bars around each point provide the associated 95% confidence interval.

effort to be low. This outside option is present in many real-world competitive settings, and it distinguishes the setting of this paper from much of the existing literature on creativity, innovation, and high-powered incentives, where agents are effectively locked in to participating.

To incorporate the outside option into the empirics, I discretize outcomes and model each submission as a choice among three options: (a) entering a tweak and remaining active (“tweak”), (b) entering an original design and remaining active (“original”), or (c) entering any design and refraining from further submissions (“abandon”). Although the precise moment that a player decides to stop investing effort is not observable, we can use inactivity as a proxy. The unit of observation thus remains an individual submission, but I now categorize designs as original or as a tweak on the basis of discrete cutoffs for similarity scores, and I identify designs that are its creator’s final submission into each contest.<sup>16</sup>

The multinomial choice framework is necessary because the trade-offs between three unordered options (tweak, original, or abandon) cannot be evaluated in a linear model. For this exercise, I classify a design as a tweak if its similarity to any earlier design by the same player is 0.7 or higher and original if its maximal similarity to previous designs by that player is 0.3 or lower.<sup>17</sup> Designs with intermediate similarity scores are omitted from the exercise, as the player’s intentions are ambiguous in the intermediate range. Each action  $a$  in this choice set is assumed to have latent utility  $u_{ijk}^a$  for

submission  $i$  by player  $j$  in contest  $k$ . I model this latent utility as a function of the player’s contemporaneous probability of winning (computed using the conditional logit estimates described in section II), the fraction of the contest transpired, the number of days remaining, and a logit error term:

$$u_{ijk}^a = \alpha^a + \beta^a \times \text{Pr}(\text{Win})_{ijk} + T_{ijk}\lambda^a + D_{ijk}\theta^a + \varepsilon_{ijk}^a,$$

where  $\text{Pr}(\text{Win})_{ijk}$  is the player’s projected probability of winning at the time submission  $ijk$  is made,  $T_{ijk}$  is fraction of the contest elapsed,  $D_{ijk}$  is the number of days remaining in the contest, and  $\varepsilon_{ijk}^a$  is distributed i.i.d. type 1 E.V. Controlling for the fraction of the contest transpired and number of days remaining is especially important here, as abandonment is mechanically more likely to be observed later in a contest.

I estimate the parameters by maximum likelihood using observed behavior. I then use the results to compute the probability that a player takes each of the three actions near the end of a contest as her chances of winning vary from 1 to 0. These probabilities are shown in figure 3. Panel A plots the probability that the player tweaks, panel B that she enters an original design, and panel C that she does either but then quits. The bars around each point span the associated 95% confidence intervals.

The probability that a player tweaks and remains active (panel A) peaks at 70% when she is a strong favorite and declines monotonically to 0 with her odds of winning. The probability that the player produces an original design (panel B) follows a distinct and highly significant inverted-U pattern, peaking at approximately a one-half odds of winning. Finally, the probability that she abandons (panel C) increases from 0 to around 80% as her odds of winning fall to 0. More flexible specifications, including higher-order polynomials in win probability, yield similar patterns, as does a model where competition is specified by the number of five-star competing designs (appendix figure E1).

<sup>16</sup>Note that this measure cannot distinguish a player who stops competing immediately after her final submission from one who waits for more information but later stops competing without additional entries. Because the end result is the same, the distinction is not critical for the purposes of this paper, as both behaviors will be influenced by information available at the time of the final submission. Anecdotally, according to some designers on this platform, it is often the case that players enter their final design knowing it is their final design and do not look back.

<sup>17</sup>Because the distribution of similarity scores is continuous in the data, there is not an obvious cutoff for defining tweaks and original designs. The results that follow are robust to alternatives such as 0.6/0.4 or 0.8/0.2.

The evidence above is an important reminder that players may stop submitting designs when competition grows severe. Taken together, the evidence reveals that incentives for creativity are greatest with balanced competition: too little, and high performers lack incentive to develop new ideas; too much, and agents stop investing effort altogether. As appendix E shows, in the data, it appears that creative effort is most attractive to high-rated players when faced off against exactly one high-rated competitor.

## VI. Mechanisms: Creativity and Risk

Why do these players respond to competition by entering more original work? In conversations with creative professionals, including the panelists hired for the exercise, several claimed that competition requires them to “be bold” or “bring the ‘wow’ factor” and that it induces them to take on more creative risk. The key assumption of this interpretation is that creativity is both riskier and higher reward than incremental changes, but whether this is true is fundamentally an empirical question.

A natural approach to answering this question might be to look at the distribution of sponsors’ ratings on original designs versus tweaks, conditioning on the rating of the tweaked design (for tweaks) or the player’s highest prior rating (for originals). Original designs after a low rating are on average higher rated than tweaks to designs with low ratings, but original designs after a high rating are on average lower rated than tweaks of top-rated designs, raising the question of why a player would deviate from her top-rated work.

The problem with this approach is that ratings are censored: it is impossible to observe improvements above the five-star rating. With this top code, original designs will necessarily appear to underperform tweaks of five-star designs, as the sponsor’s rating can only go down. The data are thus inadequate for the exercise. To get around the top code, I hired a panel of five professional graphic designers to independently assess on an extended scale all 316 designs in my sample that were rated five stars by contest sponsors, and I use the panelists’ ratings to evaluate whether creativity is in fact a high-risk, high-return activity by comparing the distribution of panelist ratings on tweaks and original designs within this subsample.

The ratings exercise and results are presented in appendix H, and here I summarize the results. The mean, median, and maximum panelist rating on original designs is on average higher than that on tweaks, with differences on the order of half of a standard deviation and significant at the 1% level. But the level of disagreement (standard deviation) among them is significantly higher as well. This evidence reinforces a possible link between creativity and risk taking suggested by research in other fields, such as social psychology and neuroscience (Dewett, 2006, who finds that a willingness to take risks is positively associated with employees’ creativity in the workplace, and Limb & Braun, 2008, who show with fMRI data that jazz pianists’ prefrontal cortex—the part of

the brain responsible for planning, decision making, and self-regulation—deactivates during improvisation).

## VII. Implications and Conclusion

Within this sample of commercial logo design competitions, I thus find that high-powered incentives have nuanced, multifaceted effects on individuals’ creative output: some competition is needed to motivate high performers to develop original, untested ideas over tweaking their earlier work, but heavy competition drives them to stop investing altogether. When the two effects are considered in tandem, the evidence indicates that the likelihood that an agent produces original work is greatest with one competitor of similar ability. The results can be rationalized by a model in which creativity is inherently risky, such that creative effort involves a choice over risk. As such, the paper ties together literatures in the social psychology of creativity and the economics of tournament competition and provides new evidence on how competition shapes the intensity and direction of individuals’ creative production.

The results have direct implications for the use of incentives as a tool for promoting creativity. The foremost lesson is that competition can motivate creativity in professional settings, provided it is balanced. In designing contracts for creative workers, managers ought to consider incentives for high-quality work relative to that of peers or colleagues, in addition to the more traditional strategy of establishing a work environment with intrinsic motivators such as freedom, flexibility, and challenge. Note that the reward need not be pecuniary; the same intuition applies when workers value recognition or status.

In practice, this “Goldilocks” level of competition may be difficult to achieve, let alone determine, and finding it would likely require experimentation with the mechanism itself by a principal in another setting. In this paper, the presence of one high-quality competitor was found to be sufficient to induce another high-quality player to produce original designs. A natural conjecture for other settings may be that a few competitors (or perhaps even one) of similar ability are enough to elicit creativity, while the presence of many such competitors would be more harmful than helpful for motivating creative output—although the precise thresholds may also depend on other features of the setting, such as the prize distribution.

The results are also relevant to innovation procurement practices, particularly as governments, private foundations, and firms increasingly contract for R&D through prizes and institutionalize prize competition.<sup>18</sup> Yet the applications are more general than R&D prizes alone. As previous discussion

<sup>18</sup>For example, the U.S. federal government now operates a platform (Challenge.gov) where agencies can seek solutions to both technical and nontechnical problems from the public, with hundreds of active competitions and prizes ranging from status only (nonpecuniary) to tens of million dollars. Similar platforms (e.g., Innocentive) are available to organizations outside the public sector. See Williams (2012) for a review of the literature on R&D prizes.

explains, the mechanism in this paper is fundamentally an RFP, a standard contracting device that firms and government agencies use to solicit designs or prototypes of new products, systems, and technologies, with a prize or production contract awarded to the preferred submission. These competitions often take place over multiple rounds, with performance scored between, much like the contests studied here.

Caution is nonetheless warranted in drawing external inference to other procurement settings, as the product being procured in this paper (a logo) is relatively simple, and proposals are heavily tailored to each client. Another potential challenge to external validity is the absence of objective evaluation criteria: the ratings and winner selection are inherently at the sponsor's subjective discretion. Yet in many RFPs, evaluation criteria similarly leave room for subjective judgments or are otherwise opaque to participants. More important, the defining feature of the R&D problem is not the ambiguity of the evaluation criteria, but rather the uncertainty around how any given product design will perform until it is tested and its performance is revealed. This uncertainty is present in all competitive R&D settings.

The final contribution is more methodological in its nature: this paper introduces new tools for measuring innovation in terms of its content. Whereas most recent attempts at content-based analysis of innovation have focused on textual analysis of patents, this paper demonstrates that even unpatentable ideas can be quantified, and it exploits a data-rich setting to study how ideas are developed and refined in response to competition. Many other questions about individual creativity and the process of innovation remain open, and this paper provides an example of how this agenda can be pursued.

#### REFERENCES

- Aghion, Philippe, Stefan Bechtold, Lea Cassar, and Holger Herz, "The Causal Effects of Competition on Innovation: Experimental Evidence," NBER working paper 19987 (2014).
- Aghion, Philippe, Nick Bloom, Richard Blundell, Rachel Griffith, and Peter Howitt, "Competition and Innovation: An Inverted-U Relationship," *Quarterly Journal of Economics* 120 (2005), 701–728.
- Amabile, Teresa M., *Creativity in Context* (Boulder, CO: Westview Press, 1996).
- Amabile, Teresa, and Mukti Khaire, "Creativity and the Role of the Leader," *Harvard Business Review* (October 2008).
- Amabile, Teresa, and Steve Kramer, "What Doesn't Motivate Creativity Can Kill It," *Harvard Business Review* (April 2012).
- Anderson, Axel, and Luis Cabral, "Go for Broke or Play It Safe? Dynamic Competition with Choice of Variance," *RAND Journal of Economics* 38 (2007), 593–609.
- Ariely, Dan, Uri Gneezy, George Loewenstein, and Nina Mazar, "Large Stakes and Big Mistakes," *Review of Economic Studies* 76 (2009), 451–469.
- Baik, Kyung Hwan, "Effort Levels in Contests with Two Asymmetric Players," *Southern Economic Journal* 61 (1994), 367–378.
- Boudreau, Kevin J., Nicola Lacetera, and Karim R. Lakhani, "Incentives and Problem Uncertainty in Innovation Contests: An Empirical Analysis," *Management Science* 57 (2011), 843–863.
- Boudreau, Kevin J., and Karim R. Lakhani, "'Open' Disclosure of Innovations, Incentives and Follow-on Reuse: Theory on Processes of Cumulative Innovation and a Field Experiment in Computational Biology," *Research Policy* 44:1 (2015).
- Boudreau, Kevin J., Karim R. Lakhani, and Michael Menietti, "Performance Responses to Competition across Skill-Levels in Rank Order Tournaments: Field Evidence and Implications for Tournament Design," *RAND Journal of Economics* 47 (2016), 140–165.
- Bradler, Christiane, Susanne Neckermann, and Arne Jonas Warne, "Incentivizing Creativity: A Large-Scale Experiment with Tournaments and Gifts," *Journal of Labor Economics* 37 (2019), 793–851.
- Brown, Jennifer, "Quitters Never Win: The (Adverse) Incentive Effects of Competing with Superstars," *Journal of Political Economy* 119 (2011), 982–1013.
- Brown, Keith C., W. V. Harlow, and Laura T. Starks, "Of Tournaments and Temptations: An Analysis of Managerial Incentives in the Mutual Fund Industry," *Journal of Finance* 51 (1996), 85–110.
- Cabral, Luis, "R&D Competition When Firms Choose Variance," *Journal of Economics and Management Strategy* 12 (2003), 139–150.
- Charness, Gary, and Daniela Grieco, "Creativity and Financial Incentives," *Journal of the European Economic Association* 17 (2019), 454–496.
- Che, Yeon-Koo, and Ian Gale, "Optimal Design of Research Contests," *American Economic Review* 93 (2003), 646–671.
- Chevalier, Judy, and Glenn Ellison, "Risk Taking by Mutual Funds as a Response to Incentives," *Journal of Political Economy* 105 (1997), 1167–1200.
- Cohen, Wesley M., "Fifty Years of Empirical Studies of Innovative Activity and Performance" (pp. 129–213), in Bronwyn Hall and Nathan Rosenberg, eds., *Handbook of the Economics of Innovation* (Amsterdam: Elsevier, 2010).
- Dewett, Todd, "Exploring the Role of Risk in Employee Creativity," *Journal of Creative Behavior* 40 (2006), 27–45.
- Ederer, Florian, "Feedback and Motivation in Dynamic Tournaments," *Journal of Economics and Management Strategy* 19 (2010), 733–769.
- Ederer, Florian, and Gustavo Manso, "Is Pay-for-Performance Detrimental to Innovation?" *Management Science* 59 (2013), 1496–1513.
- Eisenberger, Robert, and Judy Cameron, "Detrimental Effects of Reward: Reality or Myth?" *American Psychologist* 51 (1996), 1153–1166.
- , "Reward, Intrinsic Interest, and Creativity: New Findings," *American Psychologist* 53 (1998), 676–679.
- Eisenberger, Robert, and Linda Rhoades, "Incremental Effects of Reward on Creativity," *Journal of Personality and Social Psychology* 81 (2001), 728–741.
- Erat, Sanjiv, and Uri Gneezy, "Incentives for Creativity," *Experimental Economics* 19 (2016), 269–280.
- Florida, Richard, and Jim Goodnight, "Managing for Creativity," *Harvard Business Review* (July–August 2005).
- Fullerton, Richard L., and R. Preston McAfee, "Auctioning Entry into Tournaments," *Journal of Political Economy* 107 (1999), 573–605.
- Genakos, Christos, and Mario Pagliero, "Interim Rank, Risk Taking, and Performance in Dynamic Tournaments," *Journal of Political Economy* 120 (2012), 782–813.
- Gilbert, Richard, "Looking for Mr. Schumpeter: Where Are We in the Competition-Innovation Debate?" in Adam B. Jaffe, Josh Lerner, and Scott Stern, eds., *Innovation Policy and the Economy*, vol. 6 (Cambridge, MA: MIT Press, 2006).
- Gross, Daniel P., "Performance Feedback in Competitive Product Development," *RAND Journal of Economics* 48 (2017), 438–466.
- Hennessey, Beth A., and Teresa M. Amabile, "Reward, Intrinsic Motivation, and Creativity," *American Psychologist* (June 1998), 674–675.
- , "Creativity," *Annual Review of Psychology* 61 (2010), 569–598.
- Limb, Charles J., and Allen R. Braun, "Neural Substrates of Spontaneous Musical Performance: An fMRI Study of Jazz Improvisation," *PLoS One* 3:2 (2008), e1679.
- Manso, Gustavo, "Motivating Innovation," *Journal of Finance* 66 (2011), 1823–1860.
- Ross, David Gaddis, "On Evaluation Costs in Strategic Factor Markets: The Implications for Competition and Organizational Design," *Management Science* 58 (2012), 791–804.
- Schumpeter, Joseph A., *Capitalism, Socialism, and Democracy* (New York: Harper, 1942).
- Shalley, Christina E., and Greg R. Oldham, "Competition and Creative Performance: Effects of Competitor Presence and Visibility," *Creativity Research Journal* 10 (1997), 337–345.
- Shalley, Christina E., Jing Zhou, and Greg R. Oldham, "The Effects of Personal and Contextual Characteristics on Creativity: Where Should We Go from Here?" *Journal of Management* 30 (2004), 933–958.

- Sternberg, Robert J., *Cognitive Psychology*, 5th ed. (Belmont: Wadsworth, 2008).
- Taylor, Curtis R., "Digging for Golden Carrots: An Analysis of Research Tournaments," *American Economic Review* 85 (1995), 872–890.
- Terwiesch, Christian, and Yi Xu, "Innovation Contests, Open Innovation, and Multiagent Problem Solving," *Management Science* 54 (2008), 1529–1543.
- Train, Kenneth E., *Discrete Choice Methods with Simulation* (Cambridge, MA: Cambridge University Press, 2009).
- Williams, Heidi, "Innovation Inducement Prizes: Connecting Research to Policy," *Journal of Policy Analysis and Management* 31 (2012), 752–776.
- Wooten, Joel, and Karl Ulrich, "The Impact of Visibility in Innovation Tournaments: Evidence from Field Experiments," Wharton School working paper (2013).
- "Idea Generation and the Role of Feedback: Evidence from Field Experiments with Innovation Tournaments," *Production and Operations Management* 26 (2017), 80–99.