Creativity Under Fire: The Effects of Competition on Innovation and the Creative Process

Daniel P. Gross*
University of California, Berkeley

November 2, 2014

JOB MARKET PAPER

(permalink to latest draft: click here)

Abstract: Creativity is fundamental to innovation and pervasive in everyday life, yet the creative process has received only limited attention in economics and can in practice be difficult to model and measure. In this paper, I study the effect of competition on individuals' incentives for creative experimentation in the production of commercial art. Using a sample of logo design contests, and a novel, content-based measure of designs' originality, I find that competition has an inverted-U shaped effect on individuals' propensity for innovation: some competition is necessary to induce players to experiment with novel, untested ideas, but heavy competition can drive them to abandon the tournament altogether, such that experimentation is maximized by the presence of one high-quality competitor. The evidence is consistent with a generalized model of agents' choice between risky, radical innovation; more reliable, incremental innovation; and exit from a creative tournament where agents are risk-averse or face decreasing returns to improvement due to a concave success function. These results reconcile conflicting evidence from an extensive literature on the effects of competition on innovation and have direct implications for R&D policy, competition policy, and the management of organizations in creative or research industries.

JEL Classification: D82 M52 O31 O32

Keywords: Creativity; Innovation; Radical vs. incremental innovation; Competition; Creative tournament; Design contest

*Address: 530 Evans Hall, University of California-Berkeley, Berkeley, CA 94720, USA; email: grossd@econ.berkeley.edu. A draft of this paper previously circulated under the title *Creativity with Competition: Incentives for Radical versus Incremental Innovation in Creative Tournaments*. I am indebted to Ben Handel and Barry Eichengreen for their wisdom, guidance, and support in this and other projects, and to John Morgan, whose conversation and feedback provided much of the inspiration for this paper. I also thank Dominick Bartelme, Youssef Benzarti, Ernesto Dal Bó, Bo Cowgill, Jose Espin-Sanchez, Joe Farrell, Przemek Jeziorski, Jin Li, Gustavo Manso, Denis Nekipelov, Mike Powell, Matthew Rabin, Gautam Rao, Steve Tadelis, Reed Walker, Glen Weyl, Brian Wright, Noam Yuchtman, participants at the UC Berkeley IO seminar, Economics student workshop, Haas student workshop, EconCS workshop, and many others for helpful feedback and conversations at different points in this project, including the five anonymous panelists for their insights and participation. This research was supported by an NSF Graduate Research Fellowship Grant No. DGE-1106400 and an EHA Fellowship. All errors are my own.

The creative act is among the most important yet least understood phenomena in economics and the social and cognitive sciences. Technological progress – the wellspring of lasting, long-run economic growth – at its heart consists of creative solutions to familiar problems. Millions of workers in the U.S. alone are employed in creative fields ranging from research, to software development, to media and the arts,[1] and surveys show that CEOs' top concerns consistently include creativity and innovation in the firm.[2] Despite its importance to innovation, in the workplace, and in everyday life, the creative process has received only limited attention in economic research and has historically proven difficult to model and measure.

In this paper, I study the incentive effects of competition on creative production. I do so in an empirical setting where creative experimentation and competition can be both precisely measured and disentangled: commercial graphic design tournaments. Using image comparison algorithms to measure experimentation, I provide causal evidence that competition can both create and destroy incentives for innovation. I find that some competition is necessary for high-performing agents to prefer experimenting with novel, untested ideas over tweaking their earlier work, but that heavy competition discourages effort of any kind. These patterns are driven by risk-return tradeoffs inherent to innovation, which I show to be high-risk, high-return. The implication of these results is an inverted-U shaped effect of competition on innovation, with incentives for taking creative risks maximized by the presence of *one* high-quality competitor.

The challenge of motivating creativity can be naturally characterized as a principal-agent problem. Suppose a firm wants its workers to experiment with new, potentially better (lower cost or higher quality) product designs, but the firm does not observe workers' creative choices and can only reward them on the quality of their output. In this setting, failed experimentation is indistinguishable from shirking. Workers who are risk-averse or face decreasing returns to improvement, as they do in this paper, may then prefer exploiting existing solutions over experimenting if the existing method reliably yields an acceptable result – even if creative and routine effort are equally expensive. Motivating innovation will be even more difficult when creative effort is more costly than routine effort, as is often the case in practice.

To better understand the economics of the creative process, I begin by developing a model of a winner-take-all "creative tournament," building on the economics literature on innovation and tournament competition.[3]

[1]According to the U.S. Census 2010 County Business Patterns, over 15 million people are employed in the Media and Communications; Professional, Scientific, and Technical Services; Management; and Arts and Entertainment sectors alone – fields that could be considered creative professions. This total represents nearly 15 percent of U.S. employment and over 20 percent of wages, and it excludes other industries in which creativity may be valued, but not strictly essential.

[2]Annual CEO survey results from The Conference Board reveal that "stimulating innovation/creativity/enabling entrepreneurship" is consistently among chief executives' top concerns. "Innovation" was perceived as the top global challenge in 2012 and the third biggest challenge in 2013 and 2014. See http://www.conference-board.org/subsites/index.cfm?id=14514.

[3]Since the seminal contributions of Arrow (1962), countless papers have studied incentives for innovation. Wright (1983) presents an interesting theoretical comparison of patents, prizes, and research contracts as incentive mechanisms, and Scotchmer (2004) provides a summary of the literature. The model in this paper is most closely related to the work of Taylor (1995), Che and Gale (2003), Fullerton and McAfee (1999), and Terwiesch and Xu (2008). Though this mechanism is commonly described in the economics literature as an "R&D," "research," or "innovation" tournament, I refer to it in the remainder of this paper as a "creative" tournament to emphasize that it applies to creative production of all kinds – not strictly research or product development. The model in this paper also has ties to recent work on tournaments with feedback, such as Yildirim (2005) and Ederer (2010), where agents accumulate effort over multiple rounds with interim evaluation.

In the model, a principal seeks a high-value product design from a pool of workers and solicits ideas using a fixed-length tournament mechanism, awarding a prize to the preferred entry. Workers compete for the prize by entering designs in turns. At each turn, a worker must choose between experimenting with a new design, tweaking an existing design, or abandoning the tournament altogether.[4] Each submission receives immediate, public feedback on its quality, and at the end of the contest, the sponsor selects the winner. The model establishes that there is an inverted-U relationship between competition and innovation, with incentives for experimentation maximized at intermediate levels of competition.[5]

I then bring the theoretical intuition to an empirical study of design competitions similar to the model's setting, using a sample of logo design contests from a prominent online platform. In these contests, a firm ("sponsor") solicits custom designs from a community of freelance designers ("players") in exchange for a winner-take-all prize. The contests in the sample offer prizes of a few hundred dollars and on average attract nearly 35 players and 100 designs. An important feature of this setting is that the sponsor can provide interim feedback on players' designs in the form of 1- to 5-star ratings. These ratings allow players to gauge the quality of their own work and the level of competition they face. The dataset also includes the designs themselves, allowing me to study experimentation in this venue: I use image comparison algorithms similar to those used by 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 in a contest relative to prior designs by that player and her competitors.

This setting presents a unique opportunity to directly observe creative experimentation in the field. Though production of commercial advertising is interesting in its own right – advertising is a $120 billion industry in the U.S. and a $520 billion industry worldwide[6] – the design process observed here is similar to that in other settings where new products are developed. It also has parallels to the experimentation with inputs and production techniques responsible for productivity improvements in firms, including those not strictly in the business of producing cutting-edge ideas: Hendel and Spiegel (2014) study plant-level productivity at a steel mill and suggest that a large fraction of its unexplained TFP growth results from the accumulation of tweaks to its production process that are tested and implemented over time.

The sponsors' ratings are critical in this paper as a source of variation in the information that both I and the players have about the state of the competition. Using these ratings, I am able to directly estimate a

[4]The explore-exploit dilemma is endemic to a class of decision models known as bandit problems, which have received extensive coverage in the economics, statistics, and operations research literatures. Weitzman (1979) provides one of the earliest applications in economics, examining optimal stopping rules in a sequential search for an innovation. Manso (2011) and Ederer and Manso (2013) study incentives for exploration in a single-agent, dynamic two-armed bandit model and an accompanying lab experiment and find that the optimal contract for motivating innovation tolerates early failure and rewards long-term success. See Bergemann and Valimaki (2008) for other applications of bandit models in economics.

[5]These results concord with the standard result from the tournament literature that asymmetries discourage effort (e.g., Baik 1994, Brown 2011). The contribution of this paper is to embed an explore-exploit problem in the model, effectively adding a new margin along which innovation (and effort) may vary: radical versus incremental. I show that intermediate competition not only maximizes incentives to participate; it also maximizes incentives to *experiment*.

[6]Magna Global Advertising Forecast for 2014, available at http://news.magnaglobal.com/.

player's probability of winning, and the results establish that ratings are meaningful: a five-star design has 10 times the weight of a four-star design, 100 times that of a three-star design, and nearly 2,000 times that of a one-star design in the success function. Data on the time at which designs are entered by players and rated by sponsors enables me to determine what every participant knows at each point in time – and what they have yet to find out. To obtain causal estimates of the effects of feedback and competition, I exploit naturally-occurring, quasi-random variation in the timing of sponsors' ratings; the requisite assumptions are that (i) players do not act strategically on feedback they have not received, and (ii) players do not know precisely what feedback they will get until that feedback is provided. These assumptions are both intuitive and empirically supported. The empirical strategy effectively compares players' responses to feedback and competition they observe at the time of design against that which has not yet been provided.

I find that feedback and competition have large effects on creative choices. In the absence of competition, positive feedback causes players to cut back sharply on experimentation: players with the top rating enter designs that are one full standard deviation more similar to their previous work than those who have only low ratings. The effect is strongest when a player receives her first five-star rating in a contest – her next design will be a near replica of the top-rated design, on average *three* standard deviations more similar to it – and attenuates at each rung down the ratings ladder. But these effects are significantly reversed (by half or more) when high-quality competition is present. Intense competition and negative feedback also drive players to abandon a contest, as their work is unlikely to win; the probability of abandonment increases with each high-rated competitor. In both reduced-form regressions and a descriptive choice model, I find high-performers are most likely to experiment when they face exactly one high-quality competitor.

For players with poor designs, the data show that continued experimentation clearly dominates imitation of the poor-performing work. But why would a top contender ever deviate from her winning formula? The model suggests that even a top contender may wish to experiment when competition is present, *provided there is sufficient upside to experimentation.* To evaluate whether it pays to innovate, I recruit a panel of professional designers to provide independent ratings on all five-star designs in my sample and correlate their responses with these designs' originality. I find that experimentation on average results in higher-rated designs than incremental changes but that the distribution of opinion also has higher variance. These results validate one of the standard assumptions in the innovation literature – that experimentation is high-risk and high-reward – which is the necessary condition for competition to motivate innovation.

To my knowledge, this paper provides the most direct view into the creative process to-date in the economics literature. The creative act is a classic example of a black box: we can see the inputs and outputs, but we have little evidence or understanding of what happens in between. Reflecting these data constraints, empirical research has opted to measure innovation in terms of inputs (R&D spending) and outcomes (patents), when innovation is at heart about what goes on in between: individual acts of discovery and invention. Because experimentation choices cannot be inferred from R&D inputs alone, and because patent data only reveal the successes – and only the subset that are patentable and its owners are willing to disclose – we may know far less about innovation than commonly believed. This paper is an effort to fill this gap.

While creativity has only recently begun to receive attention from economists,[7] social psychologists have studied the effects of intrinsic and extrinsic motivation on creativity for decades.[8] The consensus from this literature is that creativity is inspired by intrinsic "enjoyment, interest, and personal challenge" (Hennessey and Amabile 2010), and that extrinsic pressures of reward, supervision, evaluation, and competition tend to undermine intrinsic motivation by causing workers to "feel controlled by the situation." The implication is that creativity cannot be managed: any attempts to manage creativity will backfire, and the best one can do is to provide a supportive environment for creative workers, leave them alone, and hope for the best.[9] Although intrinsic motivation is undoubtedly important to creativity, I counter these claims with evidence that individuals' creative choices respond positively to well-designed incentive schemes.[10]

The evidence that incentives for assuming creative risk are highest with moderate competition has broader implications for R&D policy, competition policy, and management of organizations in creative and research industries, which I discuss in depth in Section 6. The results also provide a partial resolution to the long-standing debate on the effects of competition on innovation, which is summarized by Gilbert (2006) and Cohen (2010). Since Schumpeter's contention that monopoly is most favorable to innovation, researchers have produced explanations for and empirical evidence of positive, negative, and inverted-U relationships between competition and innovation. The confusion results from disagreements of definition and measurement; ambiguity in the type of competition being studied; problems with econometric identification; and institutional differences, such as whether innovation is appropriable. This paper addresses these issues by establishing clear and precise measures of competition and innovation, identifying the causal effects of *information* about competition on innovation, and focusing the analysis on a setting with a fixed, winner-take-all prize and copyright protections. Moreover, as Gilbert (2006) notes, the literature has largely ignored that individuals are the source of innovation ("discoveries come from creative people"), even if patents get filed by corporations. It is precisely this gap that I seek to fill with the present paper.

[7]For examples, see Weitzman (1998) and Azoulay et al. (2011). Akcigit and Liu (2014) and Halac et al. (2014) study problems more similar to the one in this paper: Akcigit and Liu embed an explore-exploit problem into a two-player patent race, as risky and safe lines of research, and study the efficiency consequences of private information; Halac et al. study the effects of various disclosure and prize-sharing policies on effort in contests to achieve successful innovation. Charness and Grieco (2014) find that financial incentives can elicit "closed" (targeted) creativity but not "open" (blue-sky) thinking. Mokyr (1990) provides a fascinating history of technological creativity at the societal level, dating back to classical antiquity.

[8]See Hennessey and Amabile (2010) for an informative, comprehensive review of the creativity literature in psychology.

[9]As Amabile and Khaire (2008) write, "One doesn't manage creativity. One manages *for* creativity [emphasis added]."

[10]The findings of this paper are not unprecedented. A smaller, rival camp of psychologists has argued that reward can have profound effects on creativity, if only applied the right way: Eisenberger and Rhoades (2001) show in a series of experiments that creativity is enhanced by rewards when it is clear to participants that creative performance is precisely what is being rewarded. This is not to say that intrinsic motivators are unimportant or should be disregarded (e.g., as Stern (2004) shows, corporate scientists sacrifice wages for the opportunity to conduct self-directed research and publish) but rather that even creative types appreciate, and will compete for, rewards of money, status, and recognition, be it out of self-interest, a desire to share the value of one's discovery or creation with a broader audience, or both.

The paper proceeds as follows. Section 1 presents the model of winner-take-all creative competition. Section 2 introduces the empirical setting, including my approach to measuring experimentation, and describes the identification strategy. Section 3 estimates the effects of competition on creative experimentation and participation. Section 4 establishes that experimentation in this setting is high-risk, high-return, confirming the driving assumption of the model. In Section 5, I unify these results and show that experimentation is maximized with one high-quality competitor. Section 6 discusses implications of these results for policymakers, managers, and future research on innovation and the creative process. Section 7 concludes with several questions on the creative act that I believe are ripe for attention.

1 A Model of a Creative Tournament

Suppose a risk-neutral principal seeks to develop a new product design. Because R&D is risky, and designs are difficult to objectively value, the principal cannot contract directly on performance. It instead sponsors a tournament to solicit prototypes from a pool of J risk-neutral players, who enter designs in turns and receive immediate, public feedback on their quality (defined below). Each design in the competition is either generated by experimentation, which has stochastic outcomes, or incrementally adapted from the blueprints of previous entries; players who choose to continue working on a given design post-feedback can re-use the blueprint to create variants, though the original version remains in contention. At a given turn, the player must choose whether to continue participating and if so, what type of innovation to undertake: radical or incremental. At the end of the tournament, the sponsor awards a fixed, winner-take-all prize P to its favorite entry. The sponsor seeks to maximize the value of the winning design.

To hone intuition, suppose each player enters at most two designs. Let each design be characterized by latent value ν_{jt}, which only the sponsor observes (possibly sponsor-specific):

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

where j indexes players and t indexes designs. In this model, β_{jt} represents the design's quality, which may not be known ex-ante and is revealed by the sponsor's feedback. The design's value to the sponsor, ν_{jt}, is increasing and concave in its quality, and the design with the highest ν wins the contest.[11] The ε_{jt} term is an i.i.d. random shock (luck), which may arise due to idiosyncracies in the sponsor's tastes at the time a winner is chosen. Player j's probability of winning takes the following form:

[11] The decision to model designs' latent value (ν_{jt}) as a function of logged quality (β_{jt}) is taken for analytical convenience but also supported by the intuition of decreasing returns to quality. ν_{jt} could also be linear in β_{jt} and similar results would obtain: the feature of the model driving the results is the concavity of the success function.

$$Pr\,(\text{player } j \text{ wins}) = \frac{\beta_{j1}+\beta_{j2}}{\sum_{k\neq j}\left(\beta_{k1}+\beta_{k2}\right)+\beta_{j1}+\beta_{j2}} = \frac{\beta_{j1}+\beta_{j2}}{\mu_j+\beta_{j1}+\beta_{j2}} \tag{2}$$

where $\mu_j \equiv \sum_{k\neq j}\left(\beta_{k1}+\beta_{k2}\right)$ is the competition that player j faces in the contest.[12] This function is concave in the player's own quality and decreasing in the quality of competition.

Players develop and submit designs one at a time, in turns, and immediately receive public feedback that reveals β_{jt}. It is assumed that property protections are in place to prevent idea theft by competitors. Every player's first design in the contest is thus novel to that contest, and at their subsequent turn, players have three options: they can exploit (tweak, or adapt) the existing design, explore (experiment with) an entirely new design, or abandon the contest altogether. I elaborate on each of these options:

1. **Exploitation** is undertaken at cost $c > 0$ and yields a design concept of the same quality as the one being exploited. A player who chooses to exploit will tweak her first design, which has quality β_{j1}, resulting in a second-round design with $\beta_{j2} = \beta_{j1}$ and increasing her probability of winning accordingly.

   After exploitation, the player's expected probability of winning is:

$$E\left[Pr\,(\text{player } j \text{ wins} \mid \text{exploit})\right] = \frac{\beta_{j1}+\beta_{j1}}{\mu_j+\beta_{j1}+\beta_{j1}} \tag{3}$$

2. **Exploration** costs $d \geq c$ and yields either a high- or low-quality design concept, each with positive probability. Define $\alpha \geq 1$ as the exogenous degree of experimentation under this option (conversely, $\frac{1}{\alpha} \in [0,1]$ can be interpreted as the similarity of the new, experimental design and the player's first design). With probability q, experimentation will yield a high-quality design with $\beta_{j2}^H = \alpha\beta_{j1}$, and with probability $(1-q)$ it will yield a low-quality design with $\beta_{j2}^L = \frac{1}{\alpha}\beta_{j1}$. I assume $q > \frac{1}{1+\alpha}$, which implies that a design's expected quality under exploration ($E[\beta_{j2}|\text{Explore}]$) is greater than that under exploitation (β_{j1}). A risk-neutral sponsor will therefore *always* want players to explore. Note that as written, exploitation is a special case of exploration, with $\alpha = 1$.[13]

   After exploration, the player's expected probability of winning is:

$$E\left[Pr\,(\text{player } j \text{ wins} \mid \text{explore})\right] = q\cdot\left(\frac{\beta_{j1}+\beta_{j2}^H}{\mu_j+\beta_{j1}+\beta_{j2}^H}\right) + (1-q)\cdot\left(\frac{\beta_{j1}+\beta_{j2}^L}{\mu_j+\beta_{j1}+\beta_{j2}^L}\right) \tag{4}$$

[12] Note that the level of competition is determined by both the number and quality of competing designs. As Section 2 shows, a single, high-quality design can present an equal amount of competition as several lower-quality ones.

[13] In this model, I assume α and q are fixed. When α is endogenized and costless, the (risk-neutral) player's optimal α is infinite, since the experimental upside would then be unlimited and the downside bounded at zero. A natural extension to this model would be to relax experimentation costs $d(\cdot)$ and/or the probability of a successful experiment $q(\cdot)$ to vary with α. Such a model is considerably more difficult to solve and beyond the scope of this paper.

Summary of available actions and parameters in the model

	Exploit ("tweak")	Explore ("experiment")	Abandon
E[Quality]	Lower	Higher	n.a.
Risk	Safe	Risky	n.a.
Pr(Win)	Eq. (3)	Eq. (4)	Eq. (5)

3. **Abandonment** is costless: the player can always walk away. Doing so leaves the player's probability of winning unchanged, as her earlier work remains in contention.

   After abandonment, the player's probability of winning will be:

$$E\left[Pr\left(\text{player } j \text{ wins} \mid \text{abandon}\right)\right] = \frac{\beta_{j1}}{\mu_j + \beta_{j1}} \tag{5}$$

From equations 3 to 5, it is clear that feedback has three effects: it informs each player about her first design's quality, helps her improve and set expectations over her second design, and reveals the level of competition she faces. Players use this information to decide (i) whether to continue participating and (ii) whether to do so by exploring a new design or re-using a previous one, which is a choice over which kind of effort to exert: creative or rote. The model thus characterizes incentives for innovation.

In the remainder of this section, I examine a player's incentives to explore, exploit, or abandon the competition. Section 1.1 studies the conditions required for the player to prefer exploration over the alternatives and shows that these conditions lead to an inverted-U relationship between competition and innovation (proofs in Appendix B). Section 1.2 contextualizes this result in the existing literature. To simplify the mathematics, I assume the level of competition μ_j is known to player j, though the results are general to other assumptions about players' beliefs over the competition they will face, including competitors' best responses. The model can also be extended to allow players to enter an arbitrary number of designs, and the results will hold as long as players do not choose exploration for its option value.

1.1 Incentives for Radical Innovation

To simplify notation, let $F(\beta_2) = F(\beta_2|\beta_1, \mu)$ denote player j's probability of winning with a second design of quality β_2, given β_1 and μ (omitting the j subscript). The model permits four values of β_2: β_2^H, β_2^L, β_1, and 0. The first two values result from exploration, and the latter two from exploitation and abandonment, respectively. For player j to experiment on a new design, she must prefer exploration over both exploitation (incentive compatibility) and abandonment (individual rationality):

$$\underbrace{\left[qF\left(\beta_2^H\right)+(1-q)\,F\left(\beta_2^L\right)\right]\cdot P-d}_{E[\pi|\text{explore}]}>\underbrace{F\left(\beta_1\right)\cdot P-c}_{E[\pi|\text{exploit}]} \tag{IC}$$

$$\underbrace{\left[qF\left(\beta_2^H\right)+(1-q)\,F\left(\beta_2^L\right)\right]\cdot P-d}_{E[\pi|\text{explore}]}>\underbrace{F\left(0\right)\cdot P}_{E[\pi|\text{abandon}]} \tag{IR}$$

These conditions can be rearranged to be written as follows:

$$qF\left(\beta_2^H\right)+(1-q)\,F\left(\beta_2^L\right)-F\left(\beta_1\right)>\frac{d-c}{P} \tag{IC}$$

$$qF\left(\beta_2^H\right)+(1-q)\,F\left(\beta_2^L\right)-F\left(0\right)>\frac{d}{P} \tag{IR}$$

In words, the probability gains from exploration over exploitation or no action must exceed the cost differential, normalized by the prize. These conditions are less likely to be met as the cost of exploration rises, but the consideration of cost in players' decision-making is mitigated in tournaments with large prizes that dwarf experimentation costs. As written, they will generate open intervals for $\mu \in \mathbb{R}^+$ in which players will degenerately prefer one of exploration, exploitation, or abandonment. If costs were stochastic – taking a distribution, as is likely the case in practice – the conditions would similarly generate intervals in which one action is more likely than (rather than strictly preferred to) the others.

1.1.1 Exploration versus Abandonment (IR)

At what values of μ are the payoffs to exploration greatest relative to abandonment? I answer this question with the following lemma that characterizes the shape of these payoffs as a function of μ, and a proposition establishing the existence of a unique value that maximizes this function.

Lemma 1. Payoffs to exploration over abandonment. *The gains to exploration over abandonment are increasing and concave in μ when μ is small and decreasing and convex when μ is large. The gains are zero when $\mu = 0$ and approach zero from above as $\mu \longrightarrow \infty$, holding β_1 fixed.*

Proposition 1. *For all values of q, there exists a unique level of competition μ_1^* at which the gains to exploration, relative to abandonment, are maximized.*

According to Lemma 1, a player becomes likely to abandon the tournament when there is either very little competition ($\mu \ll \beta_1$) or very much competition ($\mu \gg \beta_1$). This result constitutes the first empirically testable prediction of the model. The level of competition μ_1^* at which the benefits to exploration relative to abandonment are greatest is implicitly defined by the following first-order condition:

$$q\left(\frac{-\left(1+\alpha\right)\beta_1}{\left(\left(1+\alpha\right)\beta_1+\mu_1^*\right)^2}\right)+(1-q)\left(\frac{-\left(1+\frac{1}{\alpha}\right)\beta_1}{\left(\left(1+\frac{1}{\alpha}\right)\beta_1+\mu_1^*\right)^2}\right)+\frac{\beta_1}{\left(\beta_1+\mu_1^*\right)^2}=0$$

1.1.2 Exploration versus Exploitation (IC)

I now ask the counterpart question: at what values of μ are the payoffs to exploration greatest relative to exploitation? I answer this question with a similar lemma and proposition.

Lemma 2. Payoffs to exploration over exploitation. *When $q \in (\frac{1}{1+\alpha}, \frac{1}{2})$, the gains to exploration over exploitation are decreasing and convex in μ for small μ, increasing and concave for intermediate μ, and decreasing and convex for large μ. When $q \in \left(\frac{1}{2}, \frac{3\alpha+1}{4\alpha+1}\right)$, they are increasing and convex for small μ and decreasing and convex for large μ. When $q > \frac{3\alpha+1}{4\alpha+1}$, they are increasing and concave for small μ and decreasing and convex for large μ. When $q < \frac{1}{1+\alpha}$, they are decreasing and convex for small μ and increasing and concave for large μ. In every case, the gains are zero when $\mu = 0$; when $q > \frac{1}{1+\alpha}$ ($q < \frac{1}{1+\alpha}$), they approach zero from above (below) as $\mu \longrightarrow \infty$, holding β_1 fixed.*

Proposition 2. *When $q > \frac{1}{1+\alpha}$, there exists a unique level of competition μ_2^* at which the gains to exploration, relative to exploitation, are maximized.*

Corollary. *When $q < \frac{1}{1+\alpha}$, exploration will never be preferred to exploitation.*

A player's incentive to explore over exploit depends on her relative position in the contest. Provided $q > \frac{1}{1+\alpha}$, in regions where incentive compatibility binds, a player will prefer exploration when she lags sufficiently far behind her competition, and she will prefer exploitation when she is sufficiently far ahead. These results naturally lead to a second empirical prediction: more positive feedback is expected to increase continuing players' tendency to exploit their existing work rather than experiment, but this effect will be offset by greater competition. The level of competition μ_2^* at which the benefits to exploration are maximized relative to exploitation is defined by the first-order condition for the IC constraint:

$$q\left(\frac{-\left(1+\alpha\right)\beta_1}{\left(\left(1+\alpha\right)\beta_1+\mu_2^*\right)^2}\right)+\left(1-q\right)\left(\frac{-\left(1+\frac{1}{\alpha}\right)\beta_1}{\left(\left(1+\frac{1}{\alpha}\right)\beta_1+\mu_2^*\right)^2}\right)+\frac{2\beta_1}{\left(2\beta_1+\mu_2^*\right)^2}=0$$

1.1.3 Tying it together: Exploration vs. the next-best alternative

Proposition 3. *At very low and very high μ, the IR constraint binds: the next-best alternative to exploration is abandonment. At intermediate μ, the IC constraint binds: the next-best alternative is exploitation.*

As μ increases from zero to infinity, the player's preferred action will evolve from abandonment, to exploitation, to exploration (provided in expectation it outperforms exploitation), to abandonment again. Figure 2 plots the absolute payoffs to each as the level of competition increases for an example parametrization and highlights each of these regions, holding β_1 fixed.[14] Note that the region in which players will abandon R&D due to a lack of competition is very narrow, and effectively occurs only with pure monopoly.

[Figure 2 about here]

[14] Competition in this setting is defined only relative to a player's own performance. Appendix B, Figure B.1 plots the regions where each action is preferred and the indifference curves between them in (μ, β_1)-space.

Putting the first three propositions together, the implication is an inverted-U shaped effect of competition on innovation. Provided that exploration is on average higher-quality than exploitation, there will exist an optimal, intermediate level of competition for motivating experimentation, and it will be attainable as long as experimentation costs are not so large as to make it completely infeasible for the player. This inverted-U pattern is plotted in Figure 3 for the same parametrization in Figure 2.[15]

Proposition 4. *When $q > \frac{1}{1+\alpha}$, there exists a unique level of competition $\mu^* \in [\mu_1^*, \mu_2^*]$ at which the gains to exploration are maximized relative to the player's* next-best *alternative.*

[Figure 3 about here]

The origins of this result can be traced directly to the incentive compatibility and participation constraints. Though increasing competition makes experimentation more attractive relative to incremental tweaks, doing so also reduces the gains to continued effort of any kind. At low levels of competition, incentive compatibility binds, such that greater competition increases creative effort. As competition intensifies, the participation constraint eventually binds, and further increases reduce creative effort. Incentives for creativity will generally peak at the point where the participation constraint becomes binding.

At the heart of this model is the player's choice between a gamble and a safe outcome. The concavity of the success function implies that players may prefer the certain outcome to the gamble – forgoing a positive expected quality improvement – even though they are risk-neutral. The inverted-U result is thus robust to risk-aversion, which increases the concavity of payoffs, as well as to limited risk-seeking behavior, provided the utility function does not offset the concavity of the success function.

While these results speak most directly to the incentives of the player with the last move, they carry forward to players with earlier moves. On the one hand, μ can be equally interpreted as present or anticipated, future competition. The inverted-U pattern will persist even when players internalize competitors' best responses: a player with an inordinate lead or deficit has no reason to continue, one with a solid lead can compel her competitors to abandon by exploiting, and one in a neck-and-neck state or somewhat behind will be most inclined to chance it with exploration to have a fighting chance at winning.

1.2 Remarks and Relation to Previous Literature

The inverted-U effect of competition on experimentation is intuitive. With minimal-to-no competition, the player is already assured victory and will not benefit from additional effort; with extreme competition, the gains to effort are too low to justify continued participation. These patterns are consistent with existing theoretical and empirical results from the tournament literature, which has argued that asymmetries reduce

[15] Appendix B, Figure B.2 plots the level of competition μ^* maximizing incentives for exploration against β_1, showing that players of higher standing require greater competition to induce them to experiment.

effort from both leaders and laggards (Baik 1994, Brown 2011). The contribution of this model is to consider participation jointly with the explore-exploit dilemma, which adds a new layer to the problem. At intermediate levels of competition, continued participation is justified, but experimentation may not be: with only limited competition, the player is sufficiently well-served by exploiting her previous work. Only at somewhat greater levels of competition will the player have incentives to experiment.

It is tempting to also draw comparisons against models of patent races, in which firms compete to be the first to arrive at a successful innovation, with the threshold for quality fixed and time of arrival unknown. In innovation contests such as the one modeled here, firms compete to create the highest-quality innovation prior to a deadline. Although Baye and Hoppe (2003) establish an isomorphism between the two, it requires that players are making i.i.d. draws with each experiment. A player's probability of winning in either model is then determined by the number of draws they make – their "effort." This assumption quite clearly does not carry over to the present setting, where designs are drawn from distributions varying across players and over time. Some of the intuition from patent race models nevertheless applies, such as predictions that firms that are hopelessly behind will abandon the competition (Fudenberg et al. 1983).

The model adds a simple, new explanation of an inverted-U pattern to the literature on competition and innovation, and in particular one that is distinct from that of Aghion et al. (2005), who study the effects of product market competition (PMC) on innovation. In the Aghion et al. model, industries can be technologically leveled or unleveled. In leveled industries, profits are determined by the (exogenous) degree of product market competition; in unleveled industries, a technological leader earns monopoly rents. When PMC is low, firms tend towards a leveled state, where pre-innovation rents are already high, and they therefore have little incentive to innovate. When PMC is high, one firm will live in a state of permanent technological leadership, because post-innovation rents are insufficient to motivate the laggard to catch-up innovate. Incentives for ongoing, back-and-forth innovation are therefore greatest when PMC is intermediate.

Though the Aghion et al. (2005) result is *prima facie* similar to the one in this section, it is in fact quite different in its theoretical origins. The primary point of departure is that I study R&D competition rather than product market competition – in contrast to Aghion et al., the most competitive contests will be those in which players are technologically neck-and-neck. The two models are thus complementary, in that they show that innovation responds non-monotonically to various types of competition.

2 Graphic Design Contests

I collected a randomly-drawn sample of 122 logo design contests from a widely-used online platform to study competition and the creative process.[16] 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/magazine covers, website/app mock-ups, and many others. Logo design is the modal design category on this platform and is 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 the firm's 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 non-profit organization) solicits custom designs from a community of freelance designers (players) in exchange for a fixed prize awarded to its favorite entry. The sponsor publishes a design brief describing its business, its customers, and what it 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 1- to 5-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.[17] 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 design 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.[18] 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 contention, 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.[19]

[16] The sample consists of all logo design contests with public bidding that began the week of Sept. 3-9, 2013 and every three weeks thereafter through the week of Nov. 5-11, 2013, excluding those with multiple prizes or mid-contest rule changes such as prize increases or deadline extensions. Appendix C describes the sampling procedures in greater detail.

[17] Though players can see competing designs, the site requires that all designs be original and actively enforces copyright protections. Players have numerous opportunities to report violations if they believe a design has been copied or misused in any way. 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 seem to be effective at limiting abuses.

[18] The sponsor may optionally retain the option of not awarding the prize to any entries if none are to its liking.

[19] One of the threats to identification throughout the empirical section is that the effect 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 that I do not observe. This concern is substantially mitigated by evidence from the dataset in Gross (2014), collected from the same platform, in which written feedback is occasionally made publicly available after a contest ends. In

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 both in terms of the prior feedback available (observed) at the time of submission as well as the feedback eventually provided. To account for the lags required to produce a design, I define preceding designs to be those entered at least one hour prior to a given design, and 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 dataset also includes the designs themselves. I invoke image comparison algorithms commonly used in content-based image retrieval software (similar to Google Image's Search by Image feature) to quantify the originality of each design entered into a contest relative 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 C explains exactly how they work. Each one takes a pair of digital images as inputs, summarizes them in terms of a specific, structural feature,[20] and returns a similarity index in [0,1], with a value of one indicating a perfect match and a zero indicating total dissimilarity. This index effectively measures the absolute correlation of two images' structural content.

To make this discussion more 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.[21] The first two logos have some features in common (they 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 subtle and mostly limited to the choice of font. The perceptual hash algorithm gives these logos a similarity score of 0.71, while the difference hash scores them 0.89. The appendix provides additional examples on more recognizable brands (Volkswagen and Microsoft Windows).

cases where it is observed, written feedback is only given to a small fraction of designs in a contest (on average, 12 percent), far less than are rated, and typically echoes the rating given, with statements such as "I really like this one" or "This is on the right track." This written feedback is also not disproportionately given to higher- or lower-rated designs: the frequency of each rating among designs receiving comments is approximately the same as in the data at large. Thus, although the written commentary does sometimes provide players with explicit suggestions or include expressions of (dis)taste for a particular element such as a color or font, the infrequency and irregularity with which it is provided suggests that it does not supersede the role of the 1- to 5-star ratings in practice or confound the estimation in this paper.

[20] The preferred, perceptual hash algorithm focuses on low frequency content isolated using signal processing tools, whereas the secondary, difference hash algorithm focuses on the gradient in pixel intensity. Both are robust to color inversion and changes in scale, aspect ratio, brightness, and contrast, albeit sensitive to rotation.

[21] To keep the platform from which the sample was collected anonymous, I omit identifying information.

Figure 1: Illustration of image comparison algorithms

(1) (2) (3)

Notes: 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 (2) and (3). The difference hash algorithm calculates similarity scores of 0.508 for (1) and (2) and 0.891 for (2) and (3).

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 one yields an index of originality between 0 and 1. This index is my principal measure of experimentation, and it is an empirical counterpart to the parameter $1/\alpha$ in the model. I also make use of related measures: for some specifications, I compare each design against only the best previously-rated designs by the same player or against the best previously-rated designs by competing players. Since players tend to re-use only their highest-rated work, the maximal similarity of a given design to any of that player's previous designs and maximum similarity to her highest-rated previous designs are highly correlated in practice (0.88 for the preferred algorithm, 0.87 for the alternative).

Creativity can manifest 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 intra-batch similarity as a measure of experimentation in batch work. Two designs are proximate if they are entered within 15 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. Intra-batch similarity is an alternative – and arguably better – measure of creative experimentation, reflecting players' tendency to try minor variants of the same concept or multiple concepts over a short period of time.

These measures are not without drawbacks or immune to debate. One drawback is that these algorithms require substantial dimensionality reduction and thus provide only a coarse measure of experimentation. 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. One might also question how well these algorithms emulate human perception. The examples in Figure 1 and the appendix assuage this concern; more generally, I have found these algorithms to be especially good at detecting designs that are plainly tweaks to earlier work (by my perception) versus those that are not, which is the margin that matters most for this paper. Appendix C discusses these and related issues in greater detail.

2.1 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 percent of designs are rated; less than three receive the top rating.

[Table 1 about here]

Among rated designs, and the median and modal rating is three stars (Table 2). Though fewer than four percent of rated designs receive a 5-star rating, over 40 percent of all winning designs are rated five stars, suggesting that these ratings convey substantial information about a design's quality and odds of success.[22] 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.

[Table 2 about here]

Table 3 characterizes the similarity measures used in the empirical analysis. For each design in the sample, I measure its maximal similarity to previous designs by the same player, previously-rated designs by the same player, and previously-rated designs by the player's competitors (all in the same contest). For every design batch, I calculate the maximal similarity of any two designs in that batch. Note that the analysis of intra-batch similarity is restricted to batches that are not missing any design files.

[Table 3 about here]

The designs themselves are available for 96 percent of submissions in the sample. 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.

2.1.1 Correlations of contest characteristics with outcomes

To shed light on how these contests operate and how assorted levers affect outcomes of interest, Table 4 explores the relationship of contest outcomes with prize value, feedback, and other contest characteristics. I borrow the large-sample data of Gross (2014), which uses a similar (but much larger) sample of logo design

[22] Another 33 percent of winning designs are rated 4 stars. Twenty-four percent of winning designs are unrated.

contests from the same setting to study the effects of feedback on outcomes of creative tournaments. Though the Gross (2014) dataset lacks the image files, it includes most of the other variables for these contests. As Appendix H shows, this sample is broadly similar to that of the present paper.

The specifications in columns (1) to (3) regress the number of players, designs, and designs per player (as measures of participation) on: the prize value, committed prize value, contest duration, length of the design brief, number of materials provided to be included in the design, and fraction of designs rated. In a departure from existing empirical research on tournaments, these regressions also control for the average cost of effort for all players in the contest as estimated by Gross (2014), which reflects the design difficulty and would otherwise be an omitted variable biasing the estimated effects of other variables.[23] Column (4) provides estimates from a probit model of whether sponsors of contests with uncommitted prizes choose to award the prize, implying the tournament produced a design good enough to be awarded.

[Table 4 about here]

The estimates in Table 4 suggest that an extra $100 in prize value on average attracts an additional 14.8 players, 55.4 designs, and 0.1 designs per player and increases the odds the prize will be awarded by 3.5 percent at the mean of all covariates. There is only a modest and statistically insignificant incremental effect of committed prize dollars, likely because the vast majority of uncommitted prizes in the sample are awarded anyway. Higher-cost contests have lower participation and a lower probability of being awarded. The effects of feedback are equally powerful: a sponsor who rates a high fraction of the designs in the contest will typically see fewer players enter but receive more designs from the participating players and have a much higher probability of finding a design it likes enough to award the prize. The effect of full feedback (relative to no feedback) on the probability the prize is awarded is nearly equal to that of a $300 increase in the prize – a more than doubling of the average and median prize in the sample.

2.1.2 Do ratings predict contest success? Estimating the success function

With the right data, the success function can be directly estimated. Recall from equation (1) that a design's latent value is a function of its rating and an i.i.d. extreme value error. In the data, there are five possible ratings. This latent value can thus be flexibly specified with fixed effects for each rating (or no rating). The success function can then be structurally estimated as a conditional logit model, using the observed win-lose

[23] Gross (2014) develops a semi-parametric procedure to estimate the heterogeneous cost of design for every player in every contest, under the assumption that this cost is constant for a given player in a given contest (as in the model) and the same under both exploration and exploitation (a possibility supported by the model, and a sensible approximation if the variation in cost across players is much larger than the variation in cost of each action for a given player). The procedure effectively uses players' abandonment decision to bound their contest-specific design cost, which must be less than the expected gains from their final design but greater than the gains from an additional design; these gains are estimated in course. Although the average cost in a contest is an imperfect control in that it is calculated from a selected sample of players, it nevertheless appears to be a reasonable estimate of design difficulty, for the reasons discussed in the paper.

outcomes of every design in a large sample of contests. To formalize the empirical success function, 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, ν_{ijk}, can be written as follows:

$$\nu_{ijk} = \beta_{\emptyset}\mathbb{1}(R_{ijk} = \emptyset) + \beta_1\mathbb{1}(R_{ijk} = 1) + \ldots + \beta_5\mathbb{1}(R_{ijk} = 5) + \varepsilon_{ijk} \equiv \psi_{ijk} + \varepsilon_{ijk} \tag{6}$$

This specification is closely related to the theoretical success function in equation (1), with the main difference being a restricted, discrete domain for the feedback. As in the theoretical model, the sponsor is assumed to select as winner the design with the highest value. In estimating the β parameters, each sponsor's choice set of designs is assumed to satisfy I.I.A.; in principle, the submission of a design of any rating in a given contest will reduce competing designs' chances of winning proportionally.[24] For contests with an uncommitted prize, the choice set also includes an outside option of not awarding the prize, with value normalized to zero. Letting I_{jk} be the number of designs by player j in contest k, and I_k be the total number of designs entered into that same contest k, the empirical success function for player jk takes the following form:

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

Gross (2014) estimates this model by maximum likelihood using a sample of 496,401 designs entered in 4,294 contests from the same setting. The results are reproduced in Table 5.

[Table 5 about here]

Several patterns emerge from this table. The fixed effects are precisely estimated, and the estimated value of a design is monotonically increasing in its rating. Only a 5-star design is on average preferred to the outside option. The contribution of a 5-star design to the empirical success function ($e^{1.53}$) is roughly 12 times that of a 4-star design ($e^{-0.96}$), 137 times that of a 3-star design ($e^{-3.39}$), and nearly 2,000 times that of a 1-star design ($e^{-6.02}$); competition at the top effectively only comes from other 5-star designs. As a measure of fit, the model correctly "predicts" the true winner relatively well, with the odds-on favorite winning almost half of all contests in the sample. These results demonstrate that this simple model fits the data quite well and in an intuitive way, suggesting that ratings provide considerable information about a player's probability of winning. The strong fit of the model also justifies the assumption that players can accurately assess these odds: though players do not observe the ratings on specific competing designs, they are provided with the distribution of ratings on their competitors' designs, which makes it possible for players to invoke a simple heuristic model such as the one estimated here in their decision-making.

[24] In unreported estimations, I test the I.I.A. assumption by removing certain subsets of designs from each contest – such as designs submitted as part of a batch that are not the first design in the batch – and re-estimating the model. The parameter estimates are statistically and quantitatively similar even when the choice set changes.

2.2 Empirical Methods and Identification

I exploit variation in the level and timing of the sponsor's ratings to estimate the effects of competition on players' creative choices. With timestamps on all activity in a contest, I can determine exactly what a player knows at each point in time about the sponsor's tastes for her work and the competition she faces. Identification is achieved by the quasi-random release of information: it is difficult to predict ex-ante exactly when or how often a sponsor will log onto the site to rate new entries, and even more so to predict if and when any given design will be rated. This strategy rests on two conditions: (i) that players do not know precisely what feedback they or their competitors will receive until that feedback is provided, and (ii) that players do not act strategically on feedback they have not received or cannot observe.[25]

To establish evidence that ratings are difficult to predict (conditional on the feedback provided to previous designs), I explore the relationship between feedback lags and the rating given. 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 infer the eventual ratings on their designs from the time elapsed without any feedback. Table 6 demonstrates that this is not the case. Column (1) regresses the lag in hours between the time a design is entered and the time it is rated on indicators for the rating given, restricting the sample to designs rated before the contest ends. Column (2) repeats the exercise, measuring the lag as a percent of the total contest duration. Column (3) expands the sample to all rated designs and replaces the dependent variable with an indicator for whether the design was rated prior to the contest's conclusion. I also control for the fraction of the contest elapsed at the time the design was entered, the number of previous designs by that player and her competitors, and contest and player fixed effects, and cluster standard errors by contest. Across all specifications, I find that lags in feedback provision cannot predict ratings.

[Table 6 about here]

Evidence that the second condition is met is provided in Section 3. As a first check, I estimate the effects of observed feedback on experimentation both with and without controls for unobserved/forthcoming feedback and find the results unchanged. To further confirm that this condition is satisfied, I estimate the effect of forthcoming feedback on experimentation, finding that with appropriate controls, it is indistinguishable from zero. I similarly examine players' tendency to imitate highly-rated competing designs and find no such patterns – either due to the copyright protection mechanism or, more likely, because players simply do not know which competing designs are highly rated (and thus which ones to imitate).

[25] Though this setting may seem like a natural opportunity for controlled experiments, the variation of interest is in the provision of the top rating, and this happens so infrequently in practice (only for four percent of designs in the sample; see Table 2) that it becomes infeasible to collect a controlled sample large enough to support asymptotic inference.

3 Competition and the Creative Process

The theoretical predictions can now be put to the test. Section 3.1 (as well as Appendices D and E) provides a battery of evidence that conditional on continued participation, competition induces the best-performing players to experiment more than they otherwise would. The basic estimating equation in this part of the paper is the following specification, with variants estimated throughout the analysis:

$$\begin{aligned} Similarity_{ijk} &= \beta_0 + \beta_5 \cdot \mathbb{1}(\bar{R}_{ijk} = 5) + \beta_{5c} \cdot \mathbb{1}(\bar{R}_{ijk} = 5)\mathbb{1}(\bar{R}_{-ijk} = 5) + \beta_{5p} \cdot \mathbb{1}(\bar{R}_{ijk} = 5)P_k \\ &\quad + \sum_{r=2}^{4} \beta_r \cdot \mathbb{1}(\bar{R}_{ijk} = r) + \gamma \cdot \mathbb{1}(\bar{R}_{-ijk} = 5) + \lambda DR_{ijk} + X_{ijk}\theta + \zeta_k + \varphi_j + \varepsilon_{ijk} \end{aligned}$$

where $Similarity_{ijk}$ is the maximum similarity of design ijk to any previous designs by player j in contest k; $\bar{R}_{ijk}$ is the highest rating player j has received in contest k prior to design ijk; $\bar{R}_{-ijk}$ is the highest rating player j's competitors have received prior to design ijk; P_k is the prize in contest k (measured in \$100s); DR_{ijk} is the number of days remaining in the contest at the time design ijk is entered; X_{ijk} is a vector of design-level controls; and ζ_k and φ_j are contest and player fixed effects, respectively.

It may be helpful to provide a roadmap to this part of the analysis in advance. In the first set of regressions, I estimate the specification above. In the second set, I replace the dependent variable with the similarity to that player's best, previously-rated designs, and then within-batch similarity. The third set of regressions examines the change in similarity to previously-rated designs, as a function of newly-received feedback. The fourth set of regressions tests the aforementioned identifying assumption that players are not acting on private information that I cannot observe. The fifth set of regressions tests whether players imitate high-performing competitors, which they should not be able to discern from the information they are given.

Section 3.2 provides the counterpart analysis examining the effects of competition on players' tendency to continue participating in or abandon the contest. The evidence substantiates the model's second prediction: that increasing competition can drive players to quit. The specifications in this section are similar to those of the experimentation regressions. I estimate variants of the following model:

$$\begin{aligned} Abandon_{ijk} &= \beta_0 + \sum_{r=1}^{5} \beta_r \cdot \mathbb{1}(\bar{R}_{ijk} = r) + \sum_{r=1}^{5} \gamma_r \cdot \mathbb{1}(\bar{R}_{-ijk} = r) \\ &\quad + \sum_{r=1}^{5} \delta_r \cdot \mathbb{1}(\bar{R}_{ijk} = r) N_{-ijk} \\ &\quad + \delta N_{-ijk} + \lambda DR_{ijk} + X_{ijk}\theta + \zeta_k + \varphi_j + \varepsilon_{ijk} \end{aligned}$$

where $Abandon_{ijk}$ indicates that player j entered no additional designs in contest k after design ijk; N_{-ijk} is the number of five-star designs by player j's competitors in contest k at the time of design ijk; and

$\bar{R}_{ijk}$, $\bar{R}_{-ijk}$, DR_{ijk}, X_{ijk}, ζ_k, and φ_j retain their previous definitions. The precise moment at which each player makes an active choice to abandon is impossible to measure, and I thus use inactivity as a proxy. In general, this measure does not distinguish between a "wait and see" approach that ends with abandonment versus abandonment immediately following design ijk. Since the end result is the same, the distinction is immaterial for the purposes of this paper. Note that standard errors throughout both Sections 3.1 and 3.2 are clustered by player to account for any within-player correlation in the error term.

3.1 Competition and Experimentation

3.1.1 Similarity of new designs to a player's previous designs

I begin by studying players' tendency to tweak any of their previous work in a contest. Table 7 provides estimates from regressions of the maximal similarity of each design to previous designs by the same player on indicators for the highest rating that player had previously received. All specifications include interactions of the indicator for having received the top rating with (i) the prize value (in \$100s) and (ii) a variable indicating the presence of top-rated competition, as well as contest and player fixed effects. The even-numbered columns additionally control for the fraction of the contest elapsed at the time of submission and the number of designs previously entered by the player and her competitors, which characterize the overall state and progression of the contest. Columns (3) and (4) control for future feedback on the player's earlier work; if players have contest-specific ability or other information unobserved by the researcher (e.g., sponsors' written comments), it will be accounted for by these regressions.

[Table 7 about here]

The results are consistent across all specifications in the table. Players with the top rating enter designs that are 0.3 points, or roughly one full standard deviation, more similar to previous work than players who have only low feedback (or no feedback). Roughly one third of this effect is shaved off by the presence of high-rated competition. With a highest observed rating of four stars, new designs are on average around 0.1 points more similar to previous work. This effect further attenuates as the best observed rating declines, and it is indistinguishable from zero at a best observed rating of two stars.

In practice, players tend to tweak only their highest-rated designs. Table 8, columns (1) and (2) estimate a variant on the first two columns of Table 7, regressing each design's maximal similarity to the *highest-rated* preceding designs by the same player on the same set of explanatory variables. Columns (3) and (4) use a sample of design batches and the alternative measure of experimentation: the maximal similarity of any two designs in each batch. Columns (5) and (6) repeat this latter exercise, weighting observations of batches by their size. All specifications control for contest and player fixed effects, and the table shows variants of the regressions with versus without design- and batch-level covariates.

[Table 8 about here]

The results for the design-level regressions (Columns 1 and 2) are similar to but slightly stronger than those of the previous table. Players with the top rating enter designs that are 0.35 points, or about 1.3 standard deviations, more similar to their highest-rated work in that contest, but this effect is reduced by more than half when there is top-rated competition. Players' tendency to make tweaks on their best designs is again monotonically decreasing in the highest rating they have received.

Columns (3) to (6) confirm that competition has similar effects on experimentation within batches of designs. When entering multiple designs at one time, the maximal similarity of any two designs in the batch declines 0.3 points, or approximately one standard deviation, for players with a top rating who also face top-rated competition, relative to those who do not. Top players facing competition are thus more likely to experiment not only across batches but also within them. The consistency of the results demonstrates that they are not sensitive to inclusion of controls or weighting batches by their size.

The regressions in Tables 7 and 8 use contest and player fixed effects to control for factors that are constant within contests, across players or within players, across contests, but they do not control for factors that are constant throughout a given contest for a given player, as doing so leaves too little variation for me to identify the effects of feedback and competition. Such factors may nevertheless be confounding omitted variables. For example, if players can sense their match to a particular contest, and change their behavior accordingly throughout the contest, the estimated effects may be confounded by this unobserved self-selection – though such concerns are in part relieved by the consistency of results in Table 7 controlling for forthcoming ratings. The estimates in the previous tables additionally mask potential heterogeneity that may be present in players' reactions to feedback and competition over the course of a contest.

Table 9 addresses these issues with a model in first differences. The dependent variable 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; a value of 1 indicates that the player transitioned fully from experimenting to copying; and a value of -1, the converse. The independent variables are changes in indicators for the highest rating the player has received, with the usual interactions of the top rating with the prize and the presence of top-rated competition. I estimate this model with assorted configurations of contest fixed effects, player fixed effects, and controls to account for other reasons why players' inclination to experiment may change over time, though the results are not statistically different across these specifications.

[Table 9 about here]

The results provide the most powerful evidence thus far on the effects of feedback and competition on the creative process. When a player receives her first five-star rating, her next design will be a near replica. The degree of similarity increases by nearly 0.9 points, or *three* standard deviations. Top-rated competition shaves nearly half of this effect. Given their magnitudes, these effects will be plainly visible to the naked eye (see Figure 1 for an example of what they look like in practice). The effects of a new, best rating of four-, three-, and two-stars on experimentation attenuate monotonically, similar to previous results.

Interestingly, these regressions also find that new recipients of the top rating can also be induced to experiment by larger prizes. The model of Section 1 suggests a natural explanation for this result: large prizes moderate the role of experimentation costs in players' decision-making. If experimentation is more costly (takes more time or effort) than incremental tweaks, it may only be worth doing when the prize is large. This is particularly the case for players with highly-rated work in the contest, given how the shape of and movement along a player's success function depends on the quality of her designs.

The appendix provides robustness checks and supplementary analysis. To confirm that these patterns are not an artifact of the perceptual hash algorithm, Appendix D re-estimates the regressions in the preceding tables using the difference hash algorithm to calculate similarity scores. The results are statistically and quantitatively similar. In Appendix E, I split out the effects of competition by the number of top-rated competing designs, finding no differences between the effects of one versus more than one competitor: all effects of competition on experimentation are achieved by one high-quality competitor.

This latter result is especially important for ruling out an information-based story. The fact that other designs received a 5-star rating might signal that the sponsor has diverse preferences and that experimentation has a higher likelihood of success than the player might otherwise believe. If this were the case, we should see experimentation continue to rise as more 5-star competitors pile into the contest. That this is not the case suggests that the effect is in fact the result of variation in incentives from competition.

In unreported regressions, I look for effects of five-star competition on experimentation by players with only four-star designs, and find attenuated effects that are negative but not significantly different from zero. I also explore the effect of prize commitment on experimentation, since the sponsor's outside option of not awarding the prize is itself a competing alternative – one which according to the conditional logit estimates in Table 5 is on average preferred to all but the highest-rated designs. The effect of prize commitment is not estimated to be different from zero. I similarly test for effects of the presence of four-star competition on experimentation by players with five-star designs, finding none. These results reinforce the perception that competition in effect comes from designs with the top, five-star rating.

3.1.2 Similarity of new designs 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 10 test whether players' creative choices are related to forthcoming, not-yet-available feedback. If an omitted determinant of creative choices is correlated with the feedback, then it would appear as if experimentation responds to future ratings, but if the identifying assumptions hold, I should only find zeros.

[Table 10 about here]

The specification in Column (1) regresses a design's maximal similarity to the player's best designs that will eventually be – but have not yet been – rated on indicators for the ratings they later receive. The estimates ostensibly suggest a potential failure of the identifying assumptions: although many are not significantly different from zero, the point estimates imply that players tweak these "placebo best designs" that have yet to be rated more or less depending on the rating they eventually receive, and that competition continues to induce experimentation, suggesting that it's not feedback per se that shapes creative choices, but rather some omitted factors that correlate with it. However, similarity to a high-rated placebo may in fact be the result of tweaks on an even earlier design that the placebo also happens to look like. Column (2) of the table thus controls for both the given and placebo designs' similarity to the *observed* best design at the time; Column (3) relaxes these controls to vary by the observed best rating. As a final check, I isolate the similarity to the placebo best design that cannot be explained by similarity to a 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 systematically tweak designs with positive forthcoming ratings. Feedback only relates to creative choices when it is observed at the time of design.

3.1.3 Imitation of competing designs

Though players can see a gallery of competing designs in the same contest, they see only the distribution of feedback these designs have received – not the ratings provided to specific, competing entries – and should therefore not be able to use this information to imitate highly-rated competitors. The regressions in Table 11 test this assumption by examining players' tendency to imitate competing designs. The results further illustrate the non-effects of feedback that players cannot observe.

The first two columns of the table provide estimates from regressions of similarity to the highest-rated design by competing players on indicators for its rating. As in previous specifications, the top-rating indicator is interacted with the prize and with an indicator for whether the player herself also has a top-rated design in the contest. The latter columns repeat the exercise with first-differenced variants of the same specifications.

There is little evidence in this table that players imitate highly-rated competitors in any systematic way – likely because they are simply unable to identify precisely which competitors are highly-rated. Appendix Table D.5 provides counterpart estimates using the difference hash algorithm, which suggest that if anything, players tend to deviate *away* from competitors' high-rated work.

[Table 11 about here]

3.2 Competition and Abandonment

Having established that competition induces players to experiment, it remains to be seen how competition affects players' decision to continue in versus abandon a contest. In Table 12 I examine the effect of a player's first rating and the competition she faces when it is received on the probability she subsequently enters more designs. I focus on the first rating a player receives because it will typically be ex-ante unpredictable. The specifications in the table regress this measure of abandonment on dummies for each rating the player may have received, alone and interacted with the number of top-rated competing designs, the latter as a distinct regressor, and dummies for the highest rating on competing designs at the time.

[Table 12 about here]

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

Players with poor initial feedback drop out with probability close to one. Those with high initial feedback are more likely to remain active, and enter more designs at roughly a 50 percentage-point higher rate, but competition counteracts this effect: with six top-rated competing designs in the contest, a player is likely to walk away no matter what initial feedback she receives. The effect is significant at only the 10 percent level, and thus somewhat imprecise. But it appears that by driving players to abandon the contest, heavy competition can discourage creative effort just as much as an absence thereof.[26]

I also study abandonment at points in a contest other than immediately following a player's first rating. Table 13 estimates the probability that a given design is a player's final design on the feedback and competition that was observed at the time. As previously discussed, this measure of abandonment could reflect either a simultaneous choice to abandon the project or a "wait and see" strategy that yields no further action –

[26]Note that although this outcome runs counter to the interests of the principal, it may be desirable from a social welfare perspective if the high-rated designs are unlikely to be outdone. See Appendix G for further discussion.

although according to one designer who participates on this platform, it is often the case that players will enter their final design knowing it is their final design and never look back.

[Table 13 about here]

This table again estimates three linear specifications and a logit model, with the same arrangement of fixed effects, and adding the design-level controls from earlier sections. The independent variables are analogous to those in the previous table, measured at the time the given design was submitted. In the preferred, linear specification of Column (3), I find that players with a top-rated design are more likely to subsequently enter more designs, but this effect is negated by the presence of one five-star competitor, and more than offset by multiple five-star competitors – with all effects significant at the one percent level.

4 Does it Really Pay to Innovate?

Why do the designers in these contests respond to competition by experimenting with new ideas? In conversations with creative professionals (including the panelists hired for the exercise below), many have asserted that competition means that they need to "be bold" or "bring the 'wow' factor," and that it induces them to take creative risks. Gambling on a more radical, untested idea is thus a calculated and intentional choice. The implicit assumption motivating this type of creative risk-taking both in the model and in practice is that experimentation is a high-risk, high-return endeavor – the upside to experimentation is what makes it worthwhile. This assumption is pervasive not only in research, but also in the public discourse on innovation and entrepreneurship. Whether or not it is true is ultimately an empirical question.

A natural way to answer this question in the present context is to examine the distribution of sponsors' ratings on radical versus incremental designs in the sample. To do so, I categorize designs as tweaks if they have similarity to any earlier designs by the same player of 0.7 or higher and record the rating of the design they are most similar to; I classify designs as experimental if their maximal similarity to earlier designs by that player is 0.3 or below and record the highest rating the player had previously received.[27] I then compute the distribution of sponsors' ratings on this subsample, conditioning on the rating of the tweaked design (for tweaks) or the highest rating previously given to that player (for experimentation).

Figure 4 illustrates these distributions. Although the modal rating in all cases is that of the conditioning variable, the figure demonstrates that experimentation is indeed higher variance than tweaking. Experimenting after poor feedback on average outperforms tweaks to designs with low ratings, especially considering that the top rating is orders of magnitude more valuable to a player than lower ratings (see Table 5 and

[27] A player's intentions are more ambiguous at intermediate values, which I accordingly omit from the estimation.

accompanying discussion). Yet experimentation appears to on average *underperform* tweaks of top-rated designs, raising the question of why a player would deviate from her top-rated work.

[Figure 4 about here]

The problem with this analysis is that the observed outcomes are censored: it is impossible to observe the fruits of experimentation beyond a five-star rating. With this top-code in place, exploration after a five-star design will necessarily appear to underperform exploitation – in the data, the sponsor's rating can only go down. The data are thus inadequate for evaluating the benefits to experimentation for players at the top. To circumvent the top-code, I hired a panel of professional graphic designers to independently assess all designs in my sample that were rated five stars by contest sponsors, and I look to the panelists' ratings for evidence that experimentation is in fact high-risk, high-return.

Results from a Panel of Professional Designers

To obtain independent appraisals of all 316 five-star designs in the sample, I hired five professional graphic designers at their regular rates to administer their own ratings to each design on an extended scale. These ratings were collected though a web-based application in which these designs were presented in random order and panelists were limited to 100 ratings per day. With each design, the panelist was provided the project title and client industry (as excerpted from the design brief in the source data) and instructed to objectively rate the "quality and appropriateness" of the given logo on a scale of 1 to 10.

Appendix F provides more detail on the survey procedure and shows the distribution of ratings from all five panelists. One panelist ("Rater 5") was particularly critical with his/her ratings and frequently ran up against the lower bound. The mass around the lower bound was apparent after the first day of the survey (i.e., after 100 ratings), and though I provide the results from this panelist in the appendix for the sake of disclosure, the decision was made at that time to exclude these ratings from subsequent analysis. The results are nevertheless robust to including ratings from this panelist above the lower bound.

To account for differences in the remaining panelists' austerity, I first normalize their ratings by demeaning, in essence removing rater fixed effects. For each design, I then compute summary statistics of the panelists' ratings (mean, median, maximum, and s.d.). As an alternative approach to aggregating panelists' ratings, I also calculate each design's score along the first principal component generated by a principal component analysis. Collectively, these summary statistics characterize the distribution of opinion on a given design. One way to think about them is as follows: if contest sponsors were randomly drawn from this population, then the realized rating on the design would be a random draw from this distribution.

I identify designs as tweaks or experimentation using the definitions above and then compare the level and variation in panelists' ratings on designs of each type. Table 14 provides the results. Designs classified as tweaks are typically rated below-average, while those classified as experimentation are typically above-average. These patterns manifest for the PCA composite, mean, and median panelist ratings; the difference in all three cases is on the order of around half of a standard deviation and is significant at the one percent level. The maximum rating that a design receives from any of the panelists is also greater for experimentation, with the difference significant at the one percent level. Yet so is the level of disagreement: the standard deviation across panelists' ratings on a given design is significantly greater for experimentation than for tweaks. The evidence thus appears to support the popular contention that radical innovation is both higher mean and higher variance than incremental innovation, even at the top.

[Table 14 about here]

5 When is Experimentation Most Likely?

The reduced-form results establish that while competition can motivate high performers to experiment with new ideas, too much competition will drive them out of the market altogether. How much is "too much"? Given that the full effect of competition on the degree of experimentation is achieved by a single, high-quality competitor, and that players are increasingly likely to quit as competition intensifies, it would be natural to conclude that incentives for active experimentation peak in the presence of one top-rated competitor – just enough to ensure that competition exists without further eroding the returns to effort.

To formalize an answer to this question, I estimate a choice model in which with each submission, a player selects from the three basic behaviors I observe in the data: (i) tweak and enter more designs, (ii) experiment and enter more designs, and (iii) do either and subsequently abandon the contest. To distinguish between players who are more likely to be truly giving up versus adopting a "wait and see" approach, I condition the latter case on the player's contemporaneous probability of winning, calculated using the conditional logit estimates in Table 5. This model will allow me to determine on which margin players are operating and to identify the conditions under which active experimentation is most likely. As before, I classify each design as a tweak if its similarity to any earlier design by the same player is 0.7 or higher and an experiment if its maximal similarity to earlier designs by that player is 0.3 or lower.

Each action in this choice set is assumed to have latent utility u_{ijk}^{a}, where i indexes submissions by player j in contest k. I model this latent utility as a function of the player's own ratings, her competitors' ratings, the time remaining in the contest, additional controls, and a logit error term:

$$\begin{aligned} u_{ijk}^a \quad = \quad & \beta_0^a + \sum_{r=1}^{5} \beta_r^a \cdot \mathbb{1}(\bar{R}_{ijk} = r) + \sum_{r=1}^{5} \gamma_r^a \cdot \mathbb{1}(\bar{R}_{-ijk} = r) \\ & + \delta_1^a \cdot \mathbb{1}(\bar{R}_{ijk} = 5)\mathbb{1}(N_{-ijk} = 1) \\ & + \delta_2^a \cdot \mathbb{1}(\bar{R}_{ijk} = 5)\mathbb{1}(N_{-ijk} = 2) \\ & + \delta_3^a \cdot \mathbb{1}(\bar{R}_{ijk} = 5)\mathbb{1}(N_{-ijk} \geq 3) \\ & + \lambda^a DR_{ijk} + X_{ijk}\theta^a + \varepsilon_{ijk}^a, \quad \varepsilon_{ijk}^a \sim \text{i.i.d. Type-I E.V.} \end{aligned}$$

The explanatory variables are defined as before: $\bar{R}_{ijk}$ is the highest rating player j has received in contest k prior to ijk, $\bar{R}_{-ijk}$ is the highest rating on competing designs, N_{-ijk} is the number of top-rated competing designs, DR_{ijk} is the number of days remaining in the contest, and X_{ijk} are controls.

I estimate the parameters by maximum likelihood using observed behavior. I then use the results to estimate the probability that a player with a 5-star design takes each of the three actions near the end of a contest, and to evaluate how these probabilities vary as the number of top-rated competitors increases from zero to three or more (for the case of no 5-star competitors, I assume the highest rating on any competing design is 4 stars). These probabilities are shown in Figure 5. Panel A plots the probability that the player tweaks and enters another design; Panel B, that she experiments and enters another design; and Panels C and D, that she abandons, conditional on her probability of winning at that time being 0.5 versus 0.05.[28] The bars around each point provide the associated 95 percent confidence interval.

[Figure 5 about here]

The probability that a player tweaks and remains active (Panel A) peaks at 52 percent when there are no 5-star competitors and is significantly lower with non-zero competition, with all differences significant at the one percent level. The probability that the player actively experiments (Panel B) peaks at 52 percent with one 5-star competitor and is significantly lower with zero, two, or three 5-star competitors (differences against zero and three significant at the one percent level; difference against two significant at the ten percent level). Panels C and D show that the probability of abandonment increases monotonically in the level of competition, and approaches 80 percent for players with a low probability of success.

Observed behavior thus appears to conform to the predictions of economic theory: when competition is low, players are on the margin between exploration and exploitation, whereas when competition is high, they

[28]To produce Panels C and D, I estimate a choice model that includes this probability as an explanatory variable. The results are not sensitive to this choice, which is taken in order to separate players who are competitive and those who lag far behind. In unreported estimations, I also split abandonment into "tweak and abandon" and "experiment and abandon." The exercise reveals that players rarely tweak and then abandon, and the probability of doing so is statistically invariant to competition, but the probability that players experiment and then abandon is significantly increasing in competition. This evidence is strikingly consistent with the theoretical model, which suggests that tweaking and abandoning isn't a margin where we should see much activity, and that players who abandon will be on the margin with experimentation.

straddle the margin between exploration and abandonment. The results of this execise also agree with the reduced-form evidence, in finding that high-rated players are most likely to actively experiment when they encounter precisely one highly-rated competitor. Panel C directly illustrates the inverted-U shaped effect of competition on experimentation and is an empirical counterpart to Figure 3.

6 Implications for Research, Management, and Policy

These results have direct implications for policies and programs to incentivize innovation, in both the workplace and the market. The foremost result is that the sharp incentives of prize competition can motivate creative effort in a work environment, but that doing so requires striking a delicate balance in the intensity of competition. In designing contracts for creative workers, managers would be keen to offer incentives for high-quality work relative to that of peers or colleagues, in addition to the traditional strategy of establishing a work environment with intrinsic motivators such as intellectual freedom, flexibility, and challenge. Another advantage of the tournament-style incentive structure is that it incorporates tolerance for failure by allowing players to recover from unsuccessful experimentation, which has been shown to be an important feature of contracts for motivating innovation (e.g., Manso 2011, Ederer and Manso 2013).

In practice, the 'Goldilocks' level of competition preferred by a principal may be difficult to achieve, much less determine. Finding it would likely require experimentation with the mechanism itself on the part of the principal, such as by changing the prize; subsidizing or restricting entry; or eliminating non-preferred players midway through the contest. In this paper, one high-quality competitor was found to be sufficient to induce another high-quality player to experiment, and further increases in competition have the effect of driving players away. As a rule of thumb for other settings, a good approximation may be to assume that one competitor of equal ability is enough to induce creative effort, but that having more than a few such competitors is likely more harmful than helpful for motivating creative workers.

The results also have bearing on design of public incentives for R&D, which is itself a creative endeavor, and the implementation of other policies (such as antitrust policy) undertaken with the intent of incentivizing innovation. Although this paper is fundamentally about individuals, the theoretical framework can be interpreted as firms competing in a winner-take-all market. This interpretation is not without some peril, as markets are inherently more dynamic than the model allows.[29] The results nevertheless shed light on the forces that define the relationship between competition and innovation, particularly in settings where post-innovation rents are much larger than participants' pre-existing rents.

[29] In many settings, incentives for innovation are shaped by both post- and pre-innovation rents, the latter of which are absent in this paper. The model could be extended to account for pre-innovation rents by re-defining the "prize" as the incremental rents and allowing it to vary by player. Even with this modification, many dynamic features of sequential innovation in the market would still be absent from the model. See Scotchmer (1991) and Green and Scotchmer (1995) for an introduction to the theory of sequential innovation and Igami (2013) for an empirical example.

Three concrete policy implications follow. The first is support for prize competition as a mechanism for generating innovation. While the focus of this paper is graphic design for marketing materials, it is conceivable to think that similar forces might be at work in other creative fields, including R&D. For example, Scotchmer (2004) recounts that in the 1970s, the U.S. Air Force established a system whereby rival companies vying for fighter jet contracts would build prototypes and fly them in competition to demonstrate quality, with the top performer winning the contract – a process which ultimately led to the F-16 and F-18 fighter jets. Looking further back in history, Brunt, Lerner, and Nicholas (2012) show that prize competitions put on by the Royal Agricultural Society of England in the 19th and 20th centuries had a significant impact on subsequent patenting activity in the targeted areas, and Moser and Nicholas (2013) show that prizes offered at the 1851 Crystal Palace Exhibition shaped the direction of inventive activity for years to come. These issues are particularly relevant today, as governments, private foundations, and firms commit ever larger sums to R&D prizes and institutionalize prize competition. The U.S. federal government now operates a website (http://www.challenge.gov/) where agencies can solicit innovative solutions to technical and non-technical problems from the public; as of February 2014, the site listed hundreds of open competitions with prizes ranging from status only (non-monetary) to upwards of 15 million dollars.

The second implication is an argument for monopoly and perfect competition potentially being equally harmful to market innovation. According to the most recent U.S. Horizontal Merger Guidelines (2010), "competition often spurs firms to innovate," and projected post-merger changes in the level of innovation is one of the government's criteria for evaluating mergers. The results of this paper suggest that a transition from no competition to some competition increases incentives for radical innovation over more modest, incremental improvements to existing technologies, but that the gains to innovation can decline to zero in crowded or overly competitive markets, leaving participants content with the status quo.

A final implication of the results in this paper is that contrary to the conventional wisdom that duplicated R&D efforts are wasteful,[30] simultaneous duplication may be ex-ante efficient: the competition of a horserace may induce more radical innovation, whose fruits might compensate for the deadweight loss of the duplicated effort. From a social welfare perspective, institutional policies prohibiting joint support of dueling research programs would then be doing more harm than good. This corollary requires further testing, but if true, it suggests not only a fresh look at existing research on the welfare impacts of R&D, but potentially important changes to both R&D policy and strategies for managing innovation in the firm.

[30]For example, see Jones (1995), Jones and Williams (1998), and Jones and Williams (2000).

7 Conclusion

Ingenuity undoubtedly occurs along a continuum, with some innovations being inherently more novel than others. Consider the smartphone: the original Apple iPhone was an extremely creative product at the time it was developed, while later generations and competitors have essentially only tweaked the first iPhone's design with hardware and operating system changes. Though most innovation consists of modest, incremental advances, many historically important innovations were more radical departures from the status quo. Understanding what inspires creative professionals to experiment with new and untested ideas is thus critical to public policy and management practices implemented to foster innovation.

This paper combines theory, data from a sample of commercial logo design tournaments, and new tools for measuring experimentation to show that while some competition is necessary to induce high performers to experiment with new ideas, excessive competition can equally discourage innovative effort. The results imply that there is an intermediate level of competition that maximizes incentives for innovation. In the setting of this paper, this intermediate value is exactly one high-performing competitor.

These results tie together the literatures in bandit decision models and tournament competition, and they provide what is to my knowledge the most direct evidence yet available on how incentives affect the creative process. The results also contribute to the long-standing debate on the effects of competition on innovation that dates back to Schumpeter (1942). Previous research has returned evidence of positive, negative, and inverted-U relationships between competition and innovation but generally suffers from inconsistencies and imprecision in its measures of competition and innovation, lack of econometric identification, and confusion regarding the economic mechanism at play. This paper addresses these issues by establishing clear and precise measures of key quantities, exploiting the arrival of information on the state of competition to identify its effects, and clarifying the mechanism responsible for the results. The end result is clear-cut evidence of an inverted-U effect of competition on innovation in winner-take-all markets.

Many questions and opportunities remain for future research. Most importantly, as Weitzman (1996) writes, "we need to understand, much better than we do, the act of human innovation." Is the essence of innovation the recombination of existing ideas in new forms, or the creation of something truly new? How does the answer change or confirm views on the merits of the increasingly unpopular classical liberal arts education, which exposes students to diverse views and approaches? Can diversity in teams compensate for a lack of breadth within its individual members? Another goal for future research is to better understand how the creative process unfolds, and especially how it adapts to constraints. A final question is whether successful innovation is stochastic, deterministic in research inputs, or something in between, as the answer has direct implications for how innovation is modeled or measured in other settings.

References

Aghion, Philippe, Nick Bloom, Richard Blundell, Rachel Griffith, and Peter Howitt, "Competition and Innovation: An Inverted-U Relationship," *Quarterly Journal of Economics*, 2005, *120* (2), 701–728.

Akcigit, Ufuk and Qingmin Liu, "The Role of Information in Innovation and Competition," 2014. Working Paper.

Amabile, Teresa M. and Mukti Khaire, "Creativity and the Role of the Leader," *Harvard Business Review*, 2008, *October*, 100–109.

Arrow, Kenneth J., "Economic Welfare and the Allocation of Resources for Invention," in "The Rate and Direction of Inventive Activity: Economic and Social Factors," Princeton: Princeton University Press, 1962.

Azoulay, Pierre, Joshua S. Graff Zivin, and Gustavo Manso, "Incentives and Creativity: Evidence from the Academic Life Sciences," *RAND Journal of Economics*, 2011, *42* (3), 527–554.

Baik, Kyung Hwan, "Effort Levels in Contests with Two Asymmetric Players," *Southern Economic Journal*, 1994, pp. 367–378.

Baye, Michael R. and Heidrun C. Hoppe, "The Strategic Equivalence of Rent-seeking, Innovation, and Patent-race Games," *Games and Economic Behavior*, 2003, *44* (2), 217–226.

Bénabou, Roland and Jean Tirole, "Intrinsic and Extrinsic Motivation," *Review of Economic Studies*, 2003, *70* (3), 489–520.

Bergemann, Dirk and Juuso Välimäki, "Bandit Problems," in Steven N. Durlauf and Lawrence E. Blume, eds., *The New Palgrave Dictionary of Economics*, 2 ed., Basingstoke: Palgrave Macmillan, 2008.

Brown, Jennifer, "Quitters Never Win: The (adverse) incentive effects of competing with superstars," *Journal of Political Economy*, 2011, *119* (5), 982–1013.

Brunt, Liam, Josh Lerner, and Tom Nicholas, "Inducement Prizes and Innovation," *Journal of Industrial Economics*, 2012, *60* (4), 657–696.

Charness, Gary and Daniela Grieco, "Creativity and Financial Incentives," 2014. Working Paper.

Che, Yeon-Koo and Ian Gale, "Optimal Design of Research Contests," *American Economic Review*, 2003, *93* (3), 646–671.

Cohen, Wesley M., "Fifty Years of Empirical Studies of Innovative Activity and Performance," *Handbook of the Economics of Innovation*, 2010, *1*, 129–213.

Ederer, Florian, "Feedback and Motivation in Dynamic Tournaments," *Journal of Economics & Management Strategy*, 2010, *19* (3), 733–769.

_ **and Gustavo Manso**, "Is Pay-for-Performance Detrimental to Innovation?," *Management Science*, 2013, *59* (7), 1496–1513.

Eisenberger, Robert and Linda Rhoades, "Incremental Effects of Reward on Creativity," *Journal of Personality and Social Psychology*, 2001, *81* (4), 728–741.

Fudenberg, Drew, Richard Gilbert, Joseph Stiglitz, and Jean Tirole, "Preemption, Leapfrogging, and Competition in Patent Races," *European Economic Review*, 1983, *22* (1), 3–31.

Fullerton, Richard L. and R. Preston McAfee, "Auctioning Entry into Tournaments," *Journal of Political Economy*, 1999, *107* (3), 573–605.

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, Volume 6*, Cambridge: The MIT Press, 2006.

Green, Jerry R. and Suzanne Scotchmer, "On the division of profit in sequential innovation," *RAND Journal of Economics*, 1995, *26* (1), 20–33.

Gross, Daniel P., "Interim Feedback in Creative Competition: Trading Off Participation for Quality," 2014. Working Paper.

Halac, Marina, Navin Kartik, and Qingmin Liu, "Contests for Experimentation," 2014. Working Paper.

Hendel, Igal and Yossi Spiegel, "Small Steps for Workers, a Giant Leap for Productivity," *American Economic Journal: Economic Policy*, 2014, *6* (1), 73–90.

Hennessey, Beth A. and Teresa M. Amabile, "Creativity," *Annual Review of Psychology*, 2010, *61*, 569–598.

Igami, Mitsuru, "Estimating the Innovator's Dilemma: Structural Analysis of Creative Destruction," 2013. Working Paper.

Jones, Charles I., "R&D-Based Models of Economic Growth," *Journal of Political Economy*, 1995, *103* (4), 759–784.

_ **and John C. Williams**, "Measuring the Social Return to R&D," *Quarterly Journal of Economics*, 1998, *113* (4), 1119–1135.

_ **and** _ , "Too Much of a Good Thing? The Economics of Investment in R&D," *Journal of Economic Growth*, 2000, *5* (1), 65–85.

Manso, Gustavo, "Motivating Innovation," *The Journal of Finance*, 2011, *66* (5), 1823–1860.

Mokyr, Joel, *The Lever of Riches: Technological Creativity and Economic Progress*, New York: Oxford University Press, 1990.

Moser, Petra and Tom Nicholas, "Prizes, Publicity and Patents: Non-monetary Awards as a Mechanism to Encourage Innovation," *Journal of Industrial Economics*, 2013, *61* (3), 763–788.

Schumpeter, Joseph A., *Capitalism, Socialism, and Democracy*, New York: Harper, 1942.

Scotchmer, Suzanne, "Standing on the Shoulders of Giants: Cumulative Research and the Patent Law," *Journal of Economic Perspectives*, 1991, *5* (1), 29–41.

Stern, Scott, "Do Scientists Pay to Be Scientists?," *Management Science*, 2004, *50* (6), 835–853.

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

Terwiesch, Christian and Yi Xu, "Innovation Contests, Open Innovation, and Multiagent Problem Solving," *Management Science*, 2008, *54* (9), 1529–1543.

U.S. Dept. of Justice and Federal Trade Commission, *Horizontal Merger Guidelines* 2010. Available at: http://www.justice.gov/atr/public/guidelines/hmg-2010.html.

Weitzman, Martin, "Optimal Search for the Best Alternative," *Econometrica*, 1979, *47* (3), 641–54.

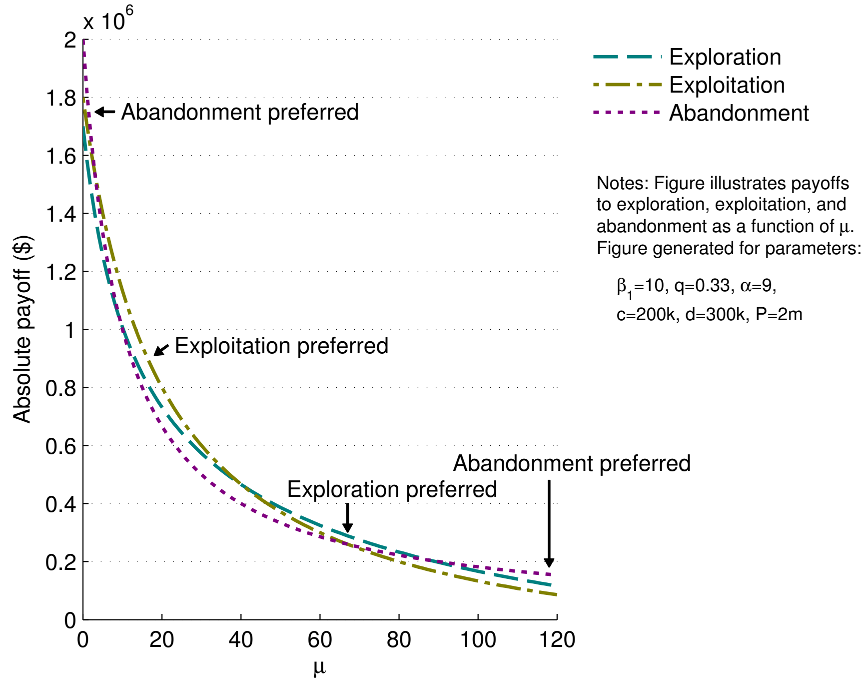
_ , "Hybridizing Growth Theory," *American Economic Review: Papers and Proceedings*, 1996, *86* (2), 207–212.

_ , "Recombinant Growth," *Quarterly Journal of Economics*, 1998, *113* (2), 331–360.

Wright, Brian D., "The Economics of Innovation Incentives: Patents, Prizes, and Research Contracts," *American Economic Review*, 1983, *73* (4), 691–707.

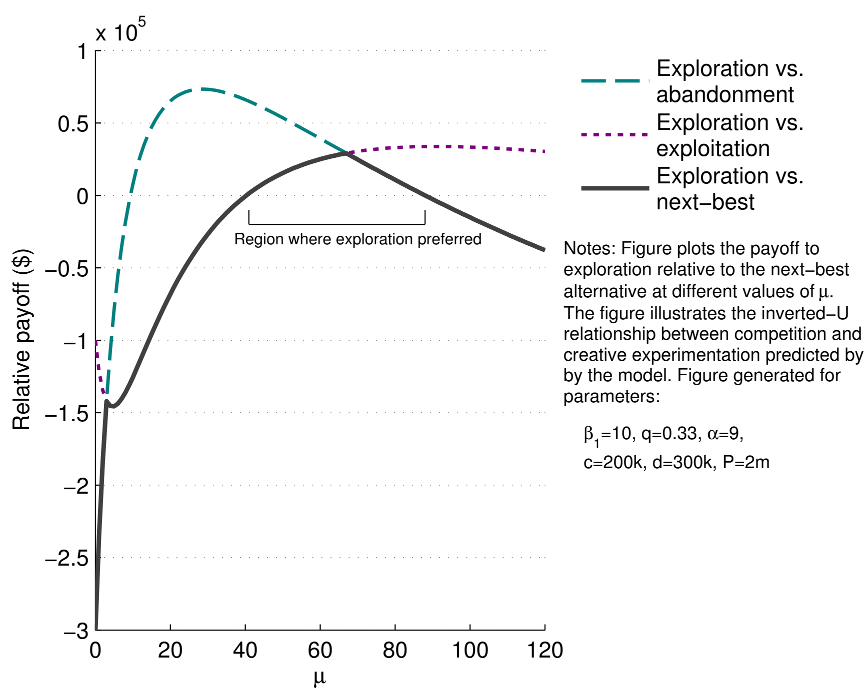
Yildirim, Huseyin, "Contests with Multiple Rounds," *Games and Economic Behavior*, 2005, *51* (1), 213–227.

Figure 2: Payoffs to each of exploration, exploitation, and dropout (example)



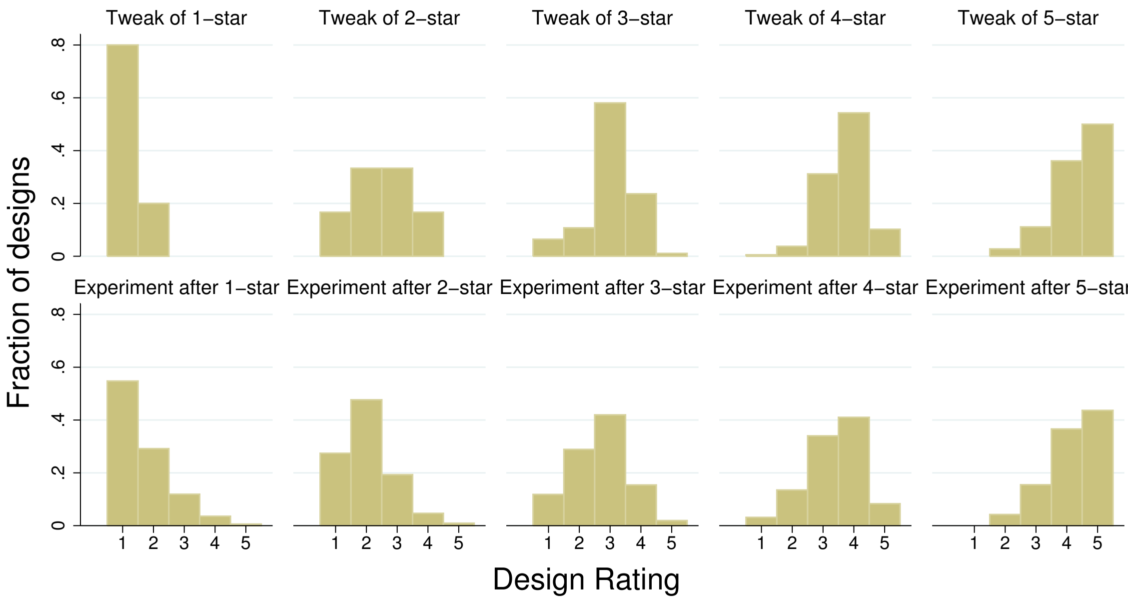
(click here to return to text from Figure 2)

Figure 3: Payoff to exploration over next-best option (example)



(click here to return to text from Figure 3)

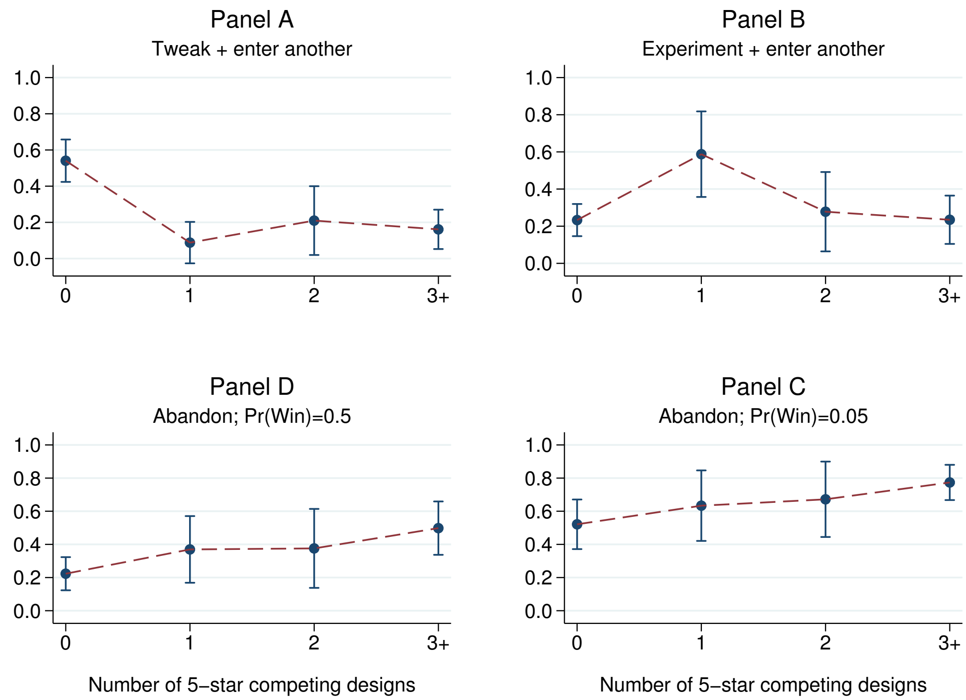
Figure 4: Sponsor ratings on tweaks vs. experimental designs



Notes: Figure shows the distribution of ratings given to tweaks versus experimental designs, grouped by the rating on the tweaked design (first row) and by the highest rating previously received by the experimenting player (second row). Each design in the sample is classified as a tweak if its maximum similarity to a previously rated design by that player is greater than 0.7 and experimentation if less than 0.3. This figure uses the perceptual hash algorithm to calculate similarity scores. Sample size in each subfigure, from left−to−right across each row: 10, 24, 93, 186, 36 (first row); 168, 321, 727, 541, 71 (second row).

(click here to return to text from Figure 4)

Figure 5: Probability of tweaking, experimenting, and abandonment as a function of competition



Notes: The figure plots the probability that a player who already has at least one 5-star rating in a contest does one of the following on (and after) a given submission: tweaks an existing design and then enters more designs (Panel A), experiments and then enters more designs (Panel B), and does either and subsequently abandons the contest, as a function of her contemporaneous probability of winning (Panels C and D). These probabilities are estimated as described in the text, and the bars around each point provide the associated 95 percent confidence interval.

The figure establishes that active experimentation is equally non-monotonic over competition in practice as it is in the theoretical model. Panel B directly illustrates this inverted-U pattern. This non-monotonicity appears to arise for the posited reasons: when competition is low, players are on the margin between tweaks and experimentation (the incentive compatibility constraint, Panels A and B); as competition increases, they are increasingly likely to stop investing, especially when their probability of winning is very low (participation constraint, Panels C and D).

(click here to return to text from Figure 5)

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
No. of players	122	33.20	24.46	19	26	39
No. 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

Notes: Table reports summary statistics for the contests in the sample. "Fraction rated" refers to the fraction of designs in each contest that gets rated. "Prize committed" indicates whether the contest prize is committed to be paid (vs. retractable). "Prize awarded" indicates whether the prize was awarded. The fraction of contests awarded awarded subsumes the fraction committed, since committed prizes are always awarded.

(click here to return to text from Table 1)

Table 2: Distribution of ratings (rated designs only)

Rating	Freq.	Percent
1 star	1,478	18.15
2 stars	2,149	26.39
3 stars	2,703	33.19
4 stars	1,498	18.39
5 stars	316	3.88
Total	**8,144**	**100**

Notes: Table tabulates rated designs, by rating. 69.3 percent (=8,144/11,758) 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 some consistency in the interpretation of ratings across contests.

(click here to return to text from Table 2)

Table 3: Similarity to preceding designs by same player and competitors, and intra-batch similarity

Using preferred algorithm: Perceptual Hash

Variable	N	Mean	SD	P10	P50	P90
Max. similarity to any of own preceding designs	5,075	0.32	0.27	0.05	0.22	0.77
Max. similarity to best of own preceding designs	3,871	0.28	0.27	0.03	0.17	0.72
Max. similarity to best of oth. preceding designs	9,709	0.14	0.1	0.04	0.13	0.27
Maximum intra-batch similarity	1,987	0.45	0.32	0.05	0.41	0.91
Image missing	11,758	0.04	0.19	0.00	0.00	0.00

Using alternative algorithm: Difference Hash

Variable	N	Mean	SD	P10	P50	P90
Max. similarity to any of own preceding designs	5,075	0.58	0.28	0.16	0.62	0.94
Max. similarity to best of own preceding designs	3,871	0.52	0.3	0.09	0.54	0.93
Max. similarity to best of oth. preceding designs	9,709	0.33	0.21	0.09	0.29	0.63
Maximum intra-batch similarity	1,987	0.69	0.28	0.23	0.77	0.98
Image missing	11,758	0.04	0.19	0.00	0.00	0.00

Notes: 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. Intra-batch similarity is calculated as the similarity of designs in a given batch to each other, where a design batch is defined to be a set of designs entered by a single player in which each design was entered within 15 minutes of another design in the set. This grouping captures players' tendency to submit multiple designs at once, which are often similar with minor variations on a theme.

(click here to return to text from Table 3)

Table 4: Correlations of contest outcomes with their characteristics

	(1) Players	(2) Designs	(3) Designs/Player	(4) Awarded
Total Prize Value ($100s)	14.828***	55.366***	0.124***	0.248***
	(0.665)	(2.527)	(0.015)	(0.042)
Committed Value ($100s)	1.860*	5.584	0.008	
	(1.118)	(4.386)	(0.025)	
Average Cost ($s)	-1.790***	-9.074***	-0.088***	-0.133***
	(0.096)	(0.353)	(0.004)	(0.010)
Fraction Rated	-14.276***	-20.056***	0.683***	0.691***
	(0.812)	(2.855)	(0.040)	(0.106)
Contest Length	0.340***	1.113***	0.003	0.007
	(0.069)	(0.251)	(0.004)	(0.010)
Words in Desc. (100s)	0.061	2.876***	0.059***	-0.158***
	(0.081)	(0.389)	(0.005)	(0.014)
Attached Materials	-0.878***	-1.557**	0.051***	-0.011
	(0.161)	(0.604)	(0.012)	(0.016)
Prize Committed	1.076	2.909	-0.023	
	(3.290)	(12.867)	(0.085)	
Constant	9.150***	-4.962	2.488***	1.967***
	(1.760)	(6.180)	(0.073)	(0.179)
N	4294	4294	4294	3298
R^2	0.63	0.65	0.31	

Notes: Table shows the estimated effect of contest attributes on overall participation and the probability that the prize is awarded from Gross (2014), controlling for the average cost of participating players. The final specification is estimated as a probit on contests without a committed prize. *, **, *** represent significance at the 0.1, 0.05, and 0.01 levels, respectively. Monthly fixed effects included but not shown. Robust SEs in parentheses.

(click here to return to text from Table 4)

Table 5: Conditional logit of win-lose outcomes on ratings

Fixed effect	**Est.**	**S.E.**	**t-stat**
Rating==5	1.53	0.07	22.17
Rating==4	-0.96	0.06	-15.35
Rating==3	-3.39	0.08	-40.01
Rating==2	-5.20	0.17	-30.16
Rating==1	-6.02	0.28	-21.82
No rating	-3.43	0.06	-55.35

Notes: Table provides estimates from conditional logit estimation of the win-lose outcome of each design as a function of its rating, using a sample of 496,401 designs (285,082 rated) in 4,294 contests from Gross (2014). Outside option is not awarding the prize, with utility normalized to zero. The design predicted by the model as the odds-on favorite wins roughly 50 percent of contests.

(click here to return to text from Table 5)

Table 6: Correlation of feedback lags with the rating given

	(1) Lag (hours)	(2) Lag (pct. of contest)	(3) Rated before end?
Rating==5	1.577	0.006	-0.023
	(3.394)	(0.016)	(0.031)
Rating==4	-2.389	-0.013*	0.009
	(1.727)	(0.007)	(0.020)
Rating==3	-0.740	-0.004	0.007
	(2.106)	(0.009)	(0.016)
Rating==2	1.167	0.005	0.006
	(1.904)	(0.008)	(0.011)
Constant	16.655*	0.134***	1.075***
	(8.514)	(0.042)	(0.050)
N	7388	7388	8144
R^2	0.45	0.48	0.63
Controls	Yes	Yes	Yes
Contest FEs	Yes	Yes	Yes
Player FEs	Yes	Yes	Yes

Notes: Table illustrates tendency for designs of different ratings to be rated more or less quickly. The results suggest that sponsors are not quicker to rate their favorite designs. Dependent variable in Column (1) is the time between submission and feedback, in hours; Column (2), this lag as a fraction of the contest length; and Column (3), an indicator for whether a design receives feedback before the contest ends. All columns control for time of entry, the number of previous designs entered by the given player and competitors, and contest and player fixed effects. *, **, *** represent significance at the 0.1, 0.05, and 0.01 levels, respectively. SEs clustered by contest in parentheses.

(click here to return to text from Table 6)

Table 7: Similarity to player's previous designs

	(1)	(2)	(3)	(4)
Player's prior best rating==5	0.284***	0.277***	0.289***	0.286***
	(0.085)	(0.087)	(0.085)	(0.087)
* 1+ competing 5-stars	-0.106*	-0.118**	-0.101*	-0.112*
	(0.058)	(0.059)	(0.058)	(0.058)
* prize value ($100s)	-0.013	-0.021	-0.011	-0.020
	(0.028)	(0.028)	(0.028)	(0.029)
Player's prior best rating==4	0.099***	0.077***	0.111***	0.090***
	(0.017)	(0.017)	(0.017)	(0.018)
Player's prior best rating==3	0.039***	0.029**	0.050***	0.039***
	(0.014)	(0.014)	(0.014)	(0.014)
Player's prior best rating==2	-0.004	-0.009	0.007	0.001
	(0.020)	(0.020)	(0.020)	(0.020)
One or more competing 5-stars	-0.014	-0.016	-0.016	-0.018
	(0.020)	(0.022)	(0.019)	(0.022)
Days remaining	-0.005*	-0.009	-0.005**	-0.009
	(0.003)	(0.007)	(0.003)	(0.007)
Constant	0.351*	0.414**	0.355*	0.419**
	(0.181)	(0.195)	(0.181)	(0.196)
N	5075	5075	5075	5075
R^2	0.47	0.47	0.47	0.47
Controls	No	Yes	No	Yes
Contest FEs	Yes	Yes	Yes	Yes
Player FEs	Yes	Yes	Yes	Yes
Forthcoming ratings	No	No	Yes	Yes

Notes: Table shows the effects of feedback on players' experimentation. Observations are designs. Dependent variable is a continuous measure of a design's maximum similarity to previous entries in the same contest by the same player, taking values in [0,1], where a value of 1 indicates the design is identical to another. The mean value of this variable in the sample is 0.32 (s.d. 0.27). Columns (2) and (4) control for time of submission and number of previous designs entered by the player and her competitors. Columns (3) and (4) additionally control for the best *forthcoming* rating on the player's not-yet-rated designs. Similarity scores in this table are calculated using a perceptual hash algorithm. Preceding designs/ratings are defined to be those entered/provided at least 60 minutes prior to the given design. *, **, *** represent significance at the 0.1, 0.05, and 0.01 levels, respectively. SEs clustered by player in parentheses.

(click here to return to text from Table 7)

Table 8: Similarity to player's best previously-rated designs & intra-batch similarity

	Designs		Batches (uwtd.)		Batches (wtd.)	
	(1)	(2)	(3)	(4)	(5)	(6)
Player's prior best rating==5	0.351***	0.362***	0.214	0.238	0.249	0.285
	(0.096)	(0.102)	(0.311)	(0.304)	(0.304)	(0.296)
* 1+ competing 5-stars	-0.204***	-0.208***	-0.302*	-0.305*	-0.300*	-0.295*
	(0.070)	(0.071)	(0.163)	(0.162)	(0.170)	(0.168)
* prize value ($100s)	-0.013	-0.018	0.016	0.015	0.010	0.009
	(0.031)	(0.033)	(0.099)	(0.097)	(0.095)	(0.093)
Player's prior best rating==4	0.119***	0.116***	0.050	0.065*	0.062*	0.086**
	(0.031)	(0.032)	(0.032)	(0.037)	(0.032)	(0.038)
Player's prior best rating==3	0.060**	0.056**	0.053	0.062*	0.051	0.065*
	(0.028)	(0.028)	(0.035)	(0.037)	(0.035)	(0.037)
Player's prior best rating==2	0.026	0.024	0.018	0.027	0.006	0.018
	(0.030)	(0.030)	(0.050)	(0.051)	(0.047)	(0.047)
One or more competing 5-stars	-0.000	0.001	0.019	0.027	0.024	0.027
	(0.022)	(0.024)	(0.048)	(0.049)	(0.052)	(0.054)
Days remaining	0.000	-0.007	0.001	-0.008	0.000	-0.005
	(0.003)	(0.008)	(0.005)	(0.011)	(0.005)	(0.011)
Constant	0.409**	0.487***	0.383***	0.507***	0.386***	0.459***
	(0.167)	(0.187)	(0.073)	(0.148)	(0.069)	(0.146)
N	3871	3871	1987	1987	1987	1987
R^2	0.53	0.53	0.57	0.57	0.58	0.58
Controls	No	Yes	No	Yes	No	Yes
Contest FEs	Yes	Yes	Yes	Yes	Yes	Yes
Player FEs	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Table shows the effects of feedback on players'experimentation. Observations in Columns (1) and (2) are designs, and 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. The mean value of this variable in the sample is 0.28 (s.d. 0.27). Observations in Columns (3) to (6) are design batches, which are defined to be a set of designs by a single player entered into a contest in close proximity (15 minutes), and dependent variable is a continuous measure of intra-batch similarity, taking values in [0,1], where a value of 1 indicates that two designs in the batch are identical. The mean value of this variable in the sample is 0.48 (s.d. 0.32). Columns (5) and (6) weight the batch regressions by batch size. All columns control for the time of submission 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 60 minutes prior to the given design. *, **, *** represent significance at the 0.1, 0.05, and 0.01 levels, respectively. SEs clustered by player in parentheses.

(click here to return to text from Table 8)

Table 9: Change in similarity to player's best previously-rated designs

	(1)	(2)	(3)	(4)	(5)	(6)
Δ(Player's best rating==5)	0.878***	0.928***	0.914***	0.885***	0.929***	0.924***
	(0.170)	(0.203)	(0.205)	(0.171)	(0.202)	(0.205)
* 1+ competing 5-stars	-0.411***	-0.419***	-0.427***	-0.414***	-0.418***	-0.429***
	(0.125)	(0.144)	(0.152)	(0.125)	(0.144)	(0.152)
* prize value ($100s)	-0.095**	-0.114**	-0.108**	-0.096**	-0.114**	-0.110**
	(0.039)	(0.049)	(0.047)	(0.040)	(0.049)	(0.048)
Δ(Player's best rating==4)	0.281***	0.268***	0.276***	0.283***	0.270***	0.279***
	(0.065)	(0.073)	(0.079)	(0.065)	(0.073)	(0.079)
Δ(Player's best rating==3)	0.150***	0.135**	0.137**	0.151***	0.136**	0.138**
	(0.058)	(0.065)	(0.069)	(0.058)	(0.065)	(0.069)
Δ(Player's best rating==2)	0.082*	0.064	0.059	0.082*	0.063	0.059
	(0.046)	(0.052)	(0.056)	(0.046)	(0.053)	(0.057)
One or more competing 5-stars	-0.004	-0.003	0.003	-0.002	-0.004	0.003
	(0.014)	(0.014)	(0.024)	(0.015)	(0.014)	(0.026)
Days remaining	-0.001	-0.001	-0.001	-0.005	-0.001	-0.004
	(0.002)	(0.002)	(0.003)	(0.004)	(0.003)	(0.007)
Constant	-0.006	-0.006	0.037	0.025	-0.015	0.060
	(0.009)	(0.008)	(0.066)	(0.040)	(0.037)	(0.077)
N	2694	2694	2694	2694	2694	2694
R^2	0.05	0.11	0.14	0.05	0.11	0.14
Controls	No	No	No	Yes	Yes	Yes
Contest FEs	Yes	No	Yes	Yes	No	Yes
Player FEs	No	Yes	Yes	No	Yes	Yes

Notes: Table shows the effects of feedback on players' experimentation. Observations are designs. 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 experimenting to copying (and a value of -1, the converse). The mean value of this variable in the sample is -0.00 (s.d. 0.23). Columns (4) to (6) control for time of submission and number of previous designs entered by the player and 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 60 minutes prior to the given design. *, **, *** represent significance at the 0.1, 0.05, and 0.01 levels, respectively. SEs clustered by player in parentheses.

(click here to return to text from Table 9)

Table 10: 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	0.227	-0.036	-0.051	-0.128
	(0.300)	(0.225)	(0.244)	(0.154)
* 1+ competing 5-stars	-0.223	-0.041	-0.052	0.012
	(0.176)	(0.114)	(0.132)	(0.099)
* prize value ($100s)	-0.029	0.005	0.010	0.028
	(0.060)	(0.053)	(0.056)	(0.042)
Player's best forthcoming rating==4	0.137*	0.006	0.003	0.000
	(0.072)	(0.068)	(0.067)	(0.061)
Player's best forthcoming rating==3	0.146***	-0.017	-0.015	-0.015
	(0.054)	(0.100)	(0.098)	(0.097)
Player's best forthcoming rating==2	0.072	-0.181*	-0.174*	-0.153
	(0.051)	(0.101)	(0.098)	(0.094)
One or more competing 5-stars	-0.079	0.002	0.005	-0.002
	(0.102)	(0.116)	(0.122)	(0.123)
Days remaining	0.011	-0.038	-0.038	-0.040
	(0.027)	(0.057)	(0.056)	(0.061)
Constant	-0.123	0.387	0.657	0.265
	(0.185)	(0.464)	(0.495)	(0.489)
N	1147	577	577	577
R^2	0.68	0.83	0.83	0.67
Controls	Yes	Yes	Yes	Yes
Contest FEs	Yes	Yes	Yes	Yes
Player FEs	Yes	Yes	Yes	Yes

Notes: Table provides a test of the effects of not-yet-available feedback on players' experimentation. Observations are designs. Dependent variable in Columns (1) to (3) is a continuous measure of a design's similarity to the best designs that the player has previously entered and 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. The mean value of this variable in the sample is 0.26 (s.d. 0.25). If players depend on sponsors' ratings for signals of quality, then forthcoming ratings should have no effect on current experimentation. The results of Column (1) suggest this may not be the case; however, similarity to an unrated design may actually be the result of both these designs being tweaks on a third design. To account for this possibility, Column (2) controls for the given design's similarity to the best previously-rated design, the best not-yet-rated design's similarity to the best previously-rated design, and their interaction. Column (3) allows these controls to vary by the best rating previously received. 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 placebo that is not explained by jointly-occurring imitation of a third design. All columns control for time of submission 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 60 minutes prior to the given design. *, **, *** represent significance at the 0.1, 0.05, and 0.01 levels, respectively. SEs clustered by player in parentheses.

(click here to return to text from Table 10)

Table 11: Similarity to competitors' best previously-rated designs

	(1)	(2)		(3)	(4)
Competitors' best==5	-0.029*	-0.003	Δ(Competitors' best==5)	-0.002	0.000
	-0.016	-0.017		-0.046	(0.046)
* 1+ own 5-stars	-0.011	-0.011	* 1+ own 5-stars	0.027	0.027
	-0.008	-0.008		-0.027	(0.027)
* prize value ($100s)	-0.005**	-0.017***	* prize value ($100s)	-0.008	-0.009
	-0.003	-0.003		-0.009	(0.009)
Competitors' best==4	0.001	0.006	Δ(Competitors' best==4)	0.035	0.035
	-0.013	-0.013		-0.033	(0.033)
Competitors' best==3	0.016	0.023*	Δ(Competitors' best==3)	0.04	0.041
	-0.012	-0.012		-0.032	(0.032)
Competitors' best==2	0.013	0.013	Δ(Competitors' best==2)	0.049	0.050
	-0.014	-0.014		-0.034	(0.034)
One or more own 5-stars	0.008	0.009	One or more own 5-stars	0.007	0.010
	-0.027	-0.028		-0.006	(0.006)
Days remaining	-0.007***	0.002	Days remaining	-0.001	0.000
	-0.001	-0.001		-0.001	(0.002)
Constant	0.059	-0.002	Constant	0.006	0.004
	-0.061	-0.063		-0.022	(0.029)
N	9709	9709	N	6065	6065
R^2	0.43	0.44	R^2	0.11	0.11
Controls	No	Yes	Controls	No	Yes
Contest FEs	Yes	Yes	Contest FEs	Yes	Yes
Player FEs	Yes	Yes	Player FEs	Yes	Yes

Notes: Table provides a test of players' ability to discern the quality of, and then imitate, competing designs. Observations are designs. Dependent variable in Columns (1) and (2) is a continuous measure of the design's similarity to the highest-rated preceding entries by *other* players, taking values in [0,1], where a value of 1 indicates that the design is identical to another. The mean value in the sample is 0.14 (s.d. 0.10). Dependent variable in Columns (3) and (4) is a continuous measure of the *change* in designs' similarity to the highest-rated preceding entries by other players, taking values in [-1,1], where a value of 0 indicates that the player's current design is equally similar to the best competing design as was her previous design, and a value of 1 indicates that the player transitioned fully from experimenting to copying (and a value of -1, the converse). The mean value of this variable in the sample is 0.00 (s.d. 0.09). In general, players are provided only the distribution of ratings on competing designs; ratings of specific competing designs are not observed. Results in this table test whether players can nevertheless identify and imitate leading competition. Columns (2) and (4) control for time of submission 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 60 minutes prior to the given design. *, **, *** represent significance at the 0.1, 0.05, and 0.01 levels, respectively. Robust SEs in parentheses.

(click here to return to text from Table 11)

Table 12: Abandonment after a player's first rating, as function of rating

	Dependent variable: Abandon after first rating			
	(1)	(2)	(3)	(4)
	Linear	Linear	Linear	Logit
Player's first rating==5	-0.352***	-0.298***	-0.481***	-1.794***
	(0.093)	(0.101)	(0.123)	(0.457)
* competing 5s	0.038**	0.024	0.076*	-0.044
	(0.019)	(0.032)	(0.042)	(0.200)
Player's first rating==4	-0.392***	-0.386***	-0.500***	-1.945***
	(0.050)	(0.068)	(0.083)	(0.267)
* competing 5s	0.020	0.035	0.056*	-0.119
	(0.015)	(0.025)	(0.031)	(0.176)
Player's first rating==3	-0.302***	-0.286***	-0.374***	-1.467***
	(0.039)	(0.057)	(0.063)	(0.212)
* competing 5s	0.013	-0.000	0.028	-0.184
	(0.013)	(0.025)	(0.030)	(0.172)
Player's first rating==2	-0.082**	-0.068	-0.135**	-0.390*
	(0.037)	(0.055)	(0.062)	(0.206)
* competing 5s	-0.016	-0.007	0.006	-0.328*
	(0.012)	(0.025)	(0.030)	(0.170)
Competitors' prior best==5	-0.040	0.030	-0.025	-0.453
	(0.082)	(0.109)	(0.130)	(0.420)
Competitors' prior best==4	0.001	0.080	0.052	-0.022
	(0.063)	(0.085)	(0.091)	(0.307)
Competitors' prior best==3	-0.079	-0.002	-0.025	-0.376
	(0.064)	(0.093)	(0.096)	(0.321)
Competitors' prior best==2	0.029	0.028	0.062	0.118
	(0.097)	(0.127)	(0.132)	(0.500)
Competitors' prior best==1	0.082	0.110	0.263	0.355
	(0.152)	(0.177)	(0.191)	(0.931)
Competing 5-star designs	0.031**	0.021	0.007	0.446**
	(0.013)	(0.024)	(0.029)	(0.178)
Days remaining	-0.019***	-0.017***	-0.026***	-0.098***
	(0.006)	(0.006)	(0.010)	(0.030)
Constant	0.899***	0.817***	0.904***	1.961***
	(0.080)	(0.095)	(0.206)	(0.720)
N	1673	1673	1673	1635
R^2	0.20	0.57	0.65	
Contest FEs	Yes	No	Yes	Yes
Player FEs	No	Yes	Yes	No

Notes: Table shows the effect of a player's first rating in a contest, and the competition at that time, on the probability that she continues to participate (by entering more designs) versus abandons the contest. Observations are contest-players. Dependent variable in all columns is an indicator for whether the player continues after her first rating [=1-$\mathbb{1}$(Abandons)]. Columns (1) to (3) estimate linear models with fixed effects; Column (4) estimates a logit model without player fixed effects, which could render the results inconsistent. *, **, *** represent significance at the 0.1, 0.05, and 0.01 levels, respectively. SEs clustered by player in parentheses.

(click here to return to text from Table 12)

Table 13: Abandonment after a given design, as function of player's ratings and competition

	Dependent variable: Abandon after given design			
	(1)	(2)	(3)	(4)
	Linear	Linear	Linear	Logit
Player's prior best rating==5	-0.017	-0.182***	-0.154***	-0.068
	(0.033)	(0.053)	(0.044)	(0.179)
* competing 5s	0.014**	0.030***	0.027***	0.068**
	(0.007)	(0.009)	(0.008)	(0.032)
Player's prior best rating==4	-0.054***	-0.104***	-0.091***	-0.212***
	(0.016)	(0.022)	(0.023)	(0.082)
* competing 5s	0.016***	0.024***	0.023***	0.091***
	(0.005)	(0.008)	(0.007)	(0.026)
Player's prior best rating==3	-0.025	-0.020	-0.006	-0.051
	(0.015)	(0.020)	(0.019)	(0.073)
* competing 5s	0.015**	0.024***	0.020**	0.074***
	(0.006)	(0.009)	(0.009)	(0.027)
Player's prior best rating==2	-0.018	0.026	0.031	-0.023
	(0.023)	(0.027)	(0.028)	(0.110)
* competing 5s	0.007	0.027*	0.026*	0.035
	(0.010)	(0.014)	(0.014)	(0.045)
Player's prior best rating==1	-0.024	0.063	0.058	-0.056
	(0.037)	(0.038)	(0.039)	(0.176)
* competing 5s	-0.024	-0.012	-0.017	-0.109
	(0.016)	(0.021)	(0.020)	(0.075)
Competitors' prior best==5	0.119***	0.085***	0.137***	0.574***
	(0.022)	(0.021)	(0.024)	(0.109)
Competitors' prior best==4	0.050***	0.020	0.065***	0.245***
	(0.017)	(0.016)	(0.017)	(0.083)
Competitors' prior best==3	-0.018	0.007	-0.006	-0.095
	(0.022)	(0.022)	(0.025)	(0.115)
Competitors' prior best==2	0.086*	0.038	0.111*	0.407*
	(0.046)	(0.048)	(0.060)	(0.219)
Competitors' prior best==1	0.013	0.093**	-0.060	0.071
	(0.053)	(0.042)	(0.054)	(0.245)
Competing 5-star designs	0.004	-0.015***	-0.003	0.020
	(0.004)	(0.004)	(0.005)	(0.020)
Days remaining	0.004	0.007**	0.009*	0.028
	(0.005)	(0.003)	(0.006)	(0.022)
Constant	0.217***	0.073**	0.061	-1.229***
	(0.044)	(0.036)	(0.102)	(0.336)
N	11758	11758	11758	11758
R^2	0.07	0.26	0.28	
Controls	Yes	Yes	Yes	Yes
Contest FEs	Yes	No	Yes	Yes
Player FEs	No	Yes	Yes	No

Notes: Table shows the effects of feedback and competition at the time a design is entered on the probability that player subsequently continues to participate (by entering more designs) versus abandons the contest. Observations are designs. Dependent variable in all columns is an indicator for whether the player continues [=1-$\mathbb{1}$(Abandons)]. Columns (1) to (3) estimate linear models with fixed effects; Column (4) estimates a logit model without player fixed effects, which could render the results inconsistent. All columns control for time of submission and number of previous designs entered by the player and her competitors. Results are qualitatively similar under a proportional hazards model. *, **, *** represent significance at the 0.1, 0.05, and 0.01 levels, respectively. SEs clustered by player in parentheses.

(click here to return to text from Table 13)

Table 14: Normalized panelist ratings on tweaks vs. experimental designs

	Outcomes under:		Diff. in
Metric	**Tweaking**	**Experimentation**	**means**
PCA score of	-0.45	0.18	0.64***
panelist ratings	(0.21)	(0.15)	p=0.008
Average rating	-0.45	0.22	0.67***
by panelists	(0.20)	(0.14)	p=0.004
Median rating	-0.46	0.23	0.69***
by panelists	(0.21)	(0.15)	p=0.005
Max rating	1.08	1.99	0.91***
by panelists	(0.22)	(0.17)	p=0.001
Disagreement (s.d.)	1.34	1.59	0.25**
among panelists	(0.10)	(0.07)	p=0.019

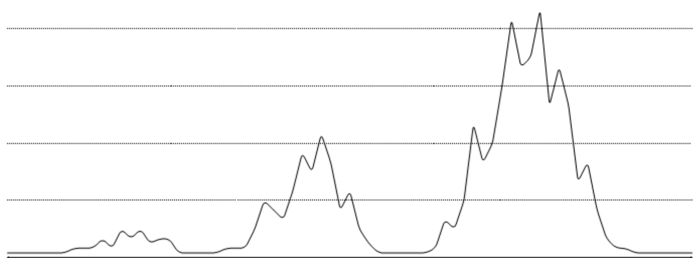
Notes: Table compares professional graphic designers' ratings on tweaks and experimental designs that received a top rating from contest sponsors. Panelists' ratings were demeaned prior to analysis. The PCA score refers to a design's score along the first component from a principal component component analysis of panelists' ratings. The other summary measures are the mean, median, max, s.d. of panelists' ratings on a given design. A design is classified as a tweak if its maximum similarity to any previous design by that player is greater than 0.7 and an experiment if it is less than 0.3. Standard errors in parentheses below each mean; results from a one-sided test of equality of means is provided to the right. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively. Similarity scores calculated using perceptual hash algorithm. Results are robust to both algorithms and alternative cutoffs for experimentation.

(click here to return to text from Table 14)

A Analogy for Search Process

A simple analogy can illustrate the nature and consequences of experimentation in this setting. Consider a sponsor with smooth preferences $f(\cdot)$ over the real line, mapped as a function over $\mathbb{R}$, as illustrated in Figure A.1. This function is ex-ante unknown to players, and perhaps even to the sponsor, but known to exist. Players begin by drawing a random number x and learning the value of $f(x)$. Since $f(\cdot)$ is smooth, the player knows that any incremental movement along $\mathbb{R}$, left or right, to x' will yield $f(x') \approx f(x)$; the outcome of a more radical deviation from x is uncertain.

Figure A.1: Sponsor's preferences over the real line



The player can use incremental search to seek out a local maximum, or the highest local maximum in a neighborhood of x, but to identify even higher local maxima or a global maximum, she will need to experiment. Experimentation from a relatively favorable initial x is likely to be high mean, but also high variance, as in Figure A.1. These are features of the model in Section 1: the expected outcome of experimentation, $q\alpha\beta_1 + (1-q)\,\frac{1}{\alpha}\beta_1$, and the difference between upside and downside outcomes, $\alpha\beta_1 - \frac{1}{\alpha}\beta_1$, are both increasing in β_1, the player's initial draw. Section 4 shows they are also features of the empirical setting.

In Figure A.1, the probability of successful experimentation after an initial draw of x is a function of $f(x)$: the better a player's existing draw, the less likely it is that experimentation will yield an even better one (and vice versa). Intuitively, when the best draw is very good, experimentation has mostly downside; when the best draw is low, it has mostly upside. This feature can naturally be incorporated into the model of Section 1 by endogenizing $q = q(\beta_1)$, with $q(\cdot)$ a decreasing, convex function of β_1 (a simple, parametric example is $q = \exp\{-\beta_1\}$). Doing so does not fundamentally change the results of Sections 1.1.1 to 1.1.3, since those results obtain from comparative statics with respect to the level of competition μ. Section 4 shows that the probability of high and low experimentation outcomes varies as expected with the initial draw.

The conclusion of Section 1 is that competition can shape incentives for experimentation when this search process is embedded in a tournament: a player with a high-quality design may be induced to experiment by the competition she faces. This result holds as long as this probability of successful experimentation is greater than the minimum threshold derived in the paper.

B Proofs of Theorems and Additional Figures

Lemma 1: The gains to exploration over abandonment are increasing and concave in μ when μ is small and decreasing and convex when μ is large. The gains are zero when $\mu = 0$ and approach zero from above as $\mu \to \infty$, holding β_1 fixed.

Proof:

Part (i): Low μ. As $\mu \longrightarrow 0$:

$$\begin{aligned} qF\left(\beta_2^H\right) + (1-q)\,F\left(\beta_2^L\right) - F(0) &= q\left(\frac{(1+\alpha)\,\beta_1}{(1+\alpha)\,\beta_1+\mu}\right) + (1-q)\left(\frac{\left(1+\frac{1}{\alpha}\right)\beta_1}{\left(1+\frac{1}{\alpha}\right)\beta_1+\mu}\right) - \left(\frac{\beta_1}{\beta_1+\mu}\right) \\ &\longrightarrow q\left(\frac{(1+\alpha)\,\beta_1}{(1+\alpha)\,\beta_1}\right) + (1-q)\left(\frac{\left(1+\frac{1}{\alpha}\right)\beta_1}{\left(1+\frac{1}{\alpha}\right)\beta_1}\right) - \left(\frac{\beta_1}{\beta_1}\right) \\ &= q + (1-q) - 1 = 0 \end{aligned}$$

$$\begin{aligned} \frac{\partial}{\partial\mu}\left[qF\left(\beta_2^H\right) + (1-q)\,F\left(\beta_2^L\right) - F(0)\right] &= q\left(\frac{-(1+\alpha)\,\beta_1}{\left((1+\alpha)\,\beta_1+\mu\right)^2}\right) + (1-q)\left(\frac{-\left(1+\frac{1}{\alpha}\right)\beta_1}{\left(\left(1+\frac{1}{\alpha}\right)\beta_1+\mu\right)^2}\right) + \frac{\beta_1}{\left(\beta_1+\mu\right)^2} \\ &\longrightarrow q\left(\frac{-1}{(1+\alpha)\,\beta_0}\right) + (1-q)\left(\frac{-1}{\left(1+\frac{1}{\alpha}\right)\beta_0}\right) + \frac{1}{\beta_0} \\ &= \frac{-q\left(1+\frac{1}{\alpha}\right) - (1-q)\,(1+\alpha) + (1+\alpha)\left(1+\frac{1}{\alpha}\right)}{(1+\alpha)\left(1+\frac{1}{\alpha}\right)\beta_1} \\ &= \frac{-\left(q+\frac{1}{\alpha}q\right) - (1-q+\alpha-\alpha q) + \left(1+\alpha+\frac{1}{\alpha}+1\right)}{(1+\alpha)\left(1+\frac{1}{\alpha}\right)\beta_1} \\ &= \frac{\alpha q - \frac{1}{\alpha}q + \alpha + \frac{1}{\alpha}}{(1+\alpha)\left(1+\frac{1}{\alpha}\right)\beta_1} = \frac{\left(\alpha^2-1\right)q + \left(1+\alpha^2\right)}{(1+\alpha)^2\,\beta_1} \longrightarrow 0^+ \end{aligned}$$

$$\frac{\partial^2}{\partial\mu^2}\left[qF\left(\beta_2^H\right) + (1-q)\,F\left(\beta_2^L\right) - F(0)\right] = q\left(\frac{2\,(1+\alpha)\,\beta_1}{\left((1+\alpha)\,\beta_1+\mu\right)^3}\right) + (1-q)\left(\frac{2\left(1+\frac{1}{\alpha}\right)\beta_1}{\left(\left(1+\frac{1}{\alpha}\right)\beta_1+\mu\right)^3}\right) - \frac{2\beta_1}{\left(\beta_1+\mu\right)^3}\ ,$$

$$\text{numerator} \longrightarrow \left(q\cdot(1+\alpha)\left(1+\tfrac{1}{\alpha}\right)^3 + (1-q)\cdot\left(1+\tfrac{1}{\alpha}\right)(1+\alpha)^3 - (1+\alpha)^3\left(1+\tfrac{1}{\alpha}\right)^3\right)2\beta_1^7 < 0$$

Part (ii): High μ. As $\mu \longrightarrow \infty$:

$$\begin{aligned}
qF\left(\beta_2^H\right) + (1-q)\,F\left(\beta_2^L\right) - F(0) &= q\left(\frac{(1+\alpha)\,\beta_1}{(1+\alpha)\,\beta_1+\mu}\right) + (1-q)\left(\frac{\left(1+\frac{1}{\alpha}\right)\beta_1}{\left(1+\frac{1}{\alpha}\right)\beta_1+\mu}\right) - \left(\frac{\beta_1}{\beta_1+\mu}\right)\\
&\longrightarrow \frac{1}{\mu}\left(q\,(1+\alpha)\,\beta_1 + (1-q)\left(1+\frac{1}{\alpha}\right)\beta_1 - \beta_1\right)\\
&= \frac{1}{\mu}\beta_1\left(\alpha q - \frac{1}{\alpha}q + \frac{1}{\alpha}\right)\\
&= \frac{1}{\alpha\mu}\beta_1\left(\left(\alpha^2-1\right)q+1\right) \longrightarrow 0^+
\end{aligned}$$

$$\begin{aligned}
\frac{\partial}{\partial\mu}\left[qF\left(\beta_2^H\right) + (1-q)\,F\left(\beta_2^L\right) - F(0)\right] &= q\left(\frac{-(1+\alpha)\,\beta_1}{((1+\alpha)\,\beta_1+\mu)^2}\right) + (1-q)\left(\frac{-\left(1+\frac{1}{\alpha}\right)\beta_1}{\left(\left(1+\frac{1}{\alpha}\right)\beta_1+\mu\right)^2}\right) + \frac{\beta_1}{(\beta_1+\mu)^2}\\
&\longrightarrow \frac{1}{\mu^2}\left(-q\,(1+\alpha)\,\beta_1 - (1-q)\left(1+\frac{1}{\alpha}\right)\beta_1 + \beta_1\right)\\
&= \frac{1}{\mu^2}\beta_1\left(-(q+\alpha q) - \left(1-q+\frac{1}{\alpha}-\frac{1}{\alpha}q\right)+1\right)\\
&= \frac{1}{\mu^2}\beta_1\left(-\alpha q + \frac{1}{\alpha}q - \frac{1}{\alpha}\right)\\
&= \frac{1}{\alpha\mu^2}\beta_1\left(-\left(\alpha^2-1\right)q-1\right) \longrightarrow 0^-
\end{aligned}$$

$$\frac{\partial^2}{\partial\mu^2}\left[qF\left(\beta_2^H\right) + (1-q)\,F\left(\beta_2^L\right) - F(0)\right] = q\left(\frac{2\,(1+\alpha)\,\beta_1}{((1+\alpha)\,\beta_1+\mu)^3}\right) + (1-q)\left(\frac{2\left(1+\frac{1}{\alpha}\right)\beta_1}{\left(\left(1+\frac{1}{\alpha}\right)\beta_1+\mu\right)^3}\right) - \frac{2\beta_1}{(\beta_1+\mu)^3}\,,$$

$$\text{numerator} \longrightarrow \left(q\cdot(1+\alpha) + (1-q)\cdot\left(1+\tfrac{1}{\alpha}\right) - 1\right)2\beta_1\mu^6 > 0$$

Taken together, these asymptotics generate a curve with the shape described.

Proposition 1: For all values of q, there exists a unique level of competition μ_1^* at which the gains to exploration, relative to abandonment, are maximized.

Proof: Existence follows from lemma and continuity of the success function. Since the difference of the success function under exploration and abandonment is quadratic in μ, it has at most two real roots, one of which is shown below to be zero, the other of which is shown to be negative. Given the shape described by the lemma, the value at which this difference is maximized must be unique.

To find the roots, set $qF\left(\beta_2^H\right) + (1-q)\,F\left(\beta_2^L\right) - F\,(0) = 0$ and solve for μ:

$$
\begin{aligned}
0 &= q\left(\frac{(1+\alpha)\,\beta_1}{(1+\alpha)\,\beta_1+\mu}\right) + (1-q)\left(\frac{\left(1+\frac{1}{\alpha}\right)\beta_1}{\left(1+\frac{1}{\alpha}\right)\beta_1+\mu}\right) - \left(\frac{\beta_1}{\beta_1+\mu}\right) \\
&= q\,(1+\alpha)\,\beta_1\left[\left(\left(1+\frac{1}{\alpha}\right)\beta_1+\mu\right)(\beta_1+\mu)\right] \\
&\quad + (1-q)\left(1+\frac{1}{\alpha}\right)\beta_1\left[\left((1+\alpha)\,\beta_1+\mu\right)(\beta_1+\mu)\right] \\
&\quad - \beta_1\left[\left((1+\alpha)\,\beta_1+\mu\right)\left(\left(1+\frac{1}{\alpha}\right)\beta_1+\mu\right)\right] \\
&= \mu^2\beta_1\left[q\,(1+\alpha) + (1-q)\left(1+\frac{1}{\alpha}\right) - 1\right] \\
&\quad + \mu\beta_1^2\left[q\left(2+\alpha+\frac{1}{\alpha}\right) + (1-q)\left(2+\alpha+\frac{1}{\alpha}\right) + q\,(1+\alpha) + (1-q)\left(1+\frac{1}{\alpha}\right) - \left(2+\alpha+\frac{1}{\alpha}\right)\right] \\
&\quad + \beta_1^3\left[q\left(2+\alpha+\frac{1}{\alpha}\right) + (1-q)\left(2+\alpha+\frac{1}{\alpha}\right) - \left(2+\alpha+\frac{1}{\alpha}\right)\right] \\
&= a\beta_1\mu^2 + b\beta_1^2\mu + c\beta_1^3\ ,
\end{aligned}
$$

where

$$
\begin{aligned}
a &= q+\alpha q+(1-q)+\frac{1}{\alpha}(1-q)-1 = \alpha q+\frac{1}{\alpha}(1-q) = \frac{1}{\alpha}\left(\left(\alpha^2-1\right)q+1\right)\ \begin{cases} >0 & \text{if } q<\frac{1}{1-\alpha^2} \\ <0 & \text{if } q>\frac{1}{1-\alpha^2} \end{cases} \\
b &= q+\alpha q+(1-q)+\frac{1}{\alpha}(1-q) = \alpha q+\frac{1}{\alpha}(1-q)+1 = \frac{1}{\alpha}(\alpha+1)\left((\alpha-1)\,q+1\right)\ \begin{cases} >0 & \text{if } q<\frac{1}{1-\alpha} \\ <0 & \text{if } q>\frac{1}{1-\alpha} \end{cases} \\
c &= 0
\end{aligned}
$$

By the quadratic formula, the roots are thus:

$$
\frac{-\left(b\beta_1^2\right)\pm\sqrt{\left(b\beta_1^2\right)^2-0}}{2\,(a\beta_1)} = \frac{-\,(b\beta_1)\pm-\,(b\beta_1)}{2a} = \begin{cases} -\beta_1\frac{b}{a} \\ 0 \end{cases}
$$

Since α is greater than one, $a<0$ and $b<0$. Thus the non-zero root is negative.

Lemma 2: When $q \in (\frac{1}{1+\alpha}, \frac{1}{2})$, the gains to exploration over exploitation are decreasing and convex in μ for small μ, increasing and concave for intermediate μ, and decreasing and convex for large μ. When $q \in \left(\frac{1}{2}, \frac{3\alpha+1}{4\alpha+1}\right)$, they are increasing and convex for small μ and decreasing and convex for large μ. When $q > \frac{3\alpha+1}{4\alpha+1}$, they are increasing and concave for small μ and decreasing and convex for large μ. When $q < \frac{1}{1+\alpha}$, they are decreasing and convex for small μ and increasing and concave for large μ. In every case, the gains are zero when $\mu = 0$; when $q > \frac{1}{1+\alpha}$ ($q < \frac{1}{1+\alpha}$), they approach zero from above (below) as $\mu \to \infty$, holding β_1 fixed.

Proof:

Part (i): Low μ. As $\mu \longrightarrow 0$:

$$
\begin{aligned}
qF\left(\beta_2^H\right) + (1-q)F\left(\beta_2^L\right) - F(\beta_1) &= q\left(\frac{(1+\alpha)\beta_1}{(1+\alpha)\beta_1+\mu}\right) + (1-q)\left(\frac{\left(1+\frac{1}{\alpha}\right)\beta_1}{\left(1+\frac{1}{\alpha}\right)\beta_1+\mu}\right) - \left(\frac{2\beta_1}{2\beta_1+\mu}\right) \\
&\longrightarrow q\left(\frac{(1+\alpha)\beta_1}{(1+\alpha)\beta_1}\right) + (1-q)\left(\frac{\left(1+\frac{1}{\alpha}\right)\beta_1}{\left(1+\frac{1}{\alpha}\right)\beta_1}\right) - \left(\frac{2\beta_1}{2\beta_1}\right) \\
&= q + (1-q) - 1 = 0
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial}{\partial\mu}\left[qF\left(\beta_2^H\right) + (1-q)F\left(\beta_2^L\right) - F(\beta_1)\right] &= q\left(\frac{-(1+\alpha)\beta_1}{((1+\alpha)\beta_1+\mu)^2}\right) + (1-q)\left(\frac{-\left(1+\frac{1}{\alpha}\right)\beta_1}{\left(\left(1+\frac{1}{\alpha}\right)\beta_1+\mu\right)^2}\right) + \frac{2\beta_1}{(2\beta_1+\mu)^2} \\
&\longrightarrow q\left(\frac{-1}{(1+\alpha)\beta_0}\right) + (1-q)\left(\frac{-1}{\left(1+\frac{1}{\alpha}\right)\beta_0}\right) + \frac{1}{2\beta_0} \\
&= \frac{-2q\left(1+\frac{1}{\alpha}\right) - 2(1-q)(1+\alpha) + (1+\alpha)\left(1+\frac{1}{\alpha}\right)}{2(1+\alpha)\left(1+\frac{1}{\alpha}\right)\beta_1} \\
&= \frac{-2\left(q+\frac{1}{\alpha}q\right) - 2(1-q+\alpha-\alpha q) + \left(1+\alpha+\frac{1}{\alpha}+1\right)}{2(1+\alpha)\left(1+\frac{1}{\alpha}\right)\beta_1} \\
&= \frac{2\alpha q - 2\frac{1}{\alpha}q - \alpha + \frac{1}{\alpha}}{2(1+\alpha)\left(1+\frac{1}{\alpha}\right)\beta_1} = \frac{\left(\alpha^2-1\right)(2q-1)}{2(1+\alpha)^2\beta_1} \longrightarrow \begin{cases} 0^+ & \text{if } q > \frac{1}{2} \\ 0^- & \text{if } q < \frac{1}{2} \end{cases}
\end{aligned}
$$

$$
\frac{\partial^2}{\partial\mu^2}\left[qF\left(\beta_2^H\right) + (1-q)F\left(\beta_2^L\right) - F(\beta_1)\right] = q\left(\frac{2(1+\alpha)\beta_1}{((1+\alpha)\beta_1+\mu)^3}\right) + (1-q)\left(\frac{2\left(1+\frac{1}{\alpha}\right)\beta_1}{\left(\left(1+\frac{1}{\alpha}\right)\beta_1+\mu\right)^3}\right) - \frac{4\beta_1}{(2\beta_1+\mu)^3},
$$

$$
\text{numerator} \longrightarrow \left(q\cdot 4(1+\alpha)\left(1+\tfrac{1}{\alpha}\right)^3 + (1-q)\cdot 4\left(1+\tfrac{1}{\alpha}\right)(1+\alpha)^3 - (1+\alpha)^3\left(1+\tfrac{1}{\alpha}\right)^3\right)4\beta_0^7 \begin{cases} > 0 & \text{if } q < \frac{3\alpha+1}{4(\alpha+1)} \\ < 0 & \text{if } q > \frac{3\alpha+1}{4(\alpha+1)} \end{cases}
$$

Part (ii): High μ. As $\mu \longrightarrow \infty$:

$$\begin{aligned}
qF\left(\beta_2^H\right) + (1-q)\,F\left(\beta_2^L\right) - F\left(\beta_1\right) &= q\left(\frac{(1+\alpha)\,\beta_1}{(1+\alpha)\,\beta_1+\mu}\right) + (1-q)\left(\frac{\left(1+\frac{1}{\alpha}\right)\beta_1}{\left(1+\frac{1}{\alpha}\right)\beta_1+\mu}\right) - \left(\frac{2\beta_1}{2\beta_1+\mu}\right) \\
&\longrightarrow \frac{1}{\mu}\left(q\,(1+\alpha)\,\beta_1 + (1-q)\left(1+\frac{1}{\alpha}\right)\beta_1 - 2\beta_1\right) \\
&= \frac{1}{\mu}\beta_1\left(\alpha q - \frac{1}{\alpha}q - 1 + \frac{1}{\alpha}\right) \\
&= \frac{1}{\alpha\mu}\beta_1\left(\left(\alpha^2-1\right)q - (\alpha-1)\right) \\
&= \frac{\alpha-1}{\alpha\mu}\beta_1\left((1+\alpha)\,q-1\right) \longrightarrow \begin{cases} 0^+ & \text{if } q > \frac{1}{1+\alpha} \\ 0^- & \text{if } q < \frac{1}{1+\alpha} \end{cases}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial}{\partial\mu}\left[qF\left(\beta_2^H\right) + (1-q)\,F\left(\beta_2^L\right) - F\left(\beta_1\right)\right] &= q\left(\frac{-(1+\alpha)\,\beta_1}{\left((1+\alpha)\,\beta_1+\mu\right)^2}\right) + (1-q)\left(\frac{-\left(1+\frac{1}{\alpha}\right)\beta_1}{\left(\left(1+\frac{1}{\alpha}\right)\beta_1+\mu\right)^2}\right) + \frac{2\beta_1}{\left(2\beta_1+\mu\right)^2} \\
&\longrightarrow \frac{1}{\mu^2}\left(-q\,(1+\alpha)\,\beta_1 - (1-q)\left(1+\frac{1}{\alpha}\right)\beta_1 + 2\beta_1\right) \\
&= \frac{1}{\mu^2}\beta_1\left(-(q+\alpha q) - \left(1-q+\frac{1}{\alpha}-\frac{1}{\alpha}q\right) + 2\right) \\
&= \frac{1}{\mu^2}\beta_1\left(-\alpha q + \frac{1}{\alpha}q + 1 - \frac{1}{\alpha}\right) \\
&= \frac{1}{\alpha\mu^2}\beta_1\left(-\left(\alpha^2-1\right)q + (\alpha-1)\right) \\
&= \frac{\alpha-1}{\alpha\mu^2}\beta_1\left(-(1+\alpha)\,q+1\right)\beta_1 \longrightarrow \begin{cases} 0^+ & \text{if } q < \frac{1}{1+\alpha} \\ 0^- & \text{if } q > \frac{1}{1+\alpha} \end{cases}
\end{aligned}$$

$$\frac{\partial^2}{\partial\mu^2}\left[qF\left(\beta_2^H\right) + (1-q)\,F\left(\beta_2^L\right) - F\left(\beta_1\right)\right] = q\left(\frac{2\,(1+\alpha)\,\beta_1}{\left((1+\alpha)\,\beta_1+\mu\right)^3}\right) + (1-q)\left(\frac{2\left(1+\frac{1}{\alpha}\right)\beta_1}{\left(\left(1+\frac{1}{\alpha}\right)\beta_1+\mu\right)^3}\right) - \frac{4\beta_1}{\left(2\beta_1+\mu\right)^3}\ ,$$

$$\text{numerator} \longrightarrow \left(q\cdot(1+\alpha) + (1-q)\cdot\left(1+\tfrac{1}{\alpha}\right) - 2\right)2\beta_1\mu^6 \ \begin{cases} >0 & \text{if } q > \frac{1}{1+\alpha} \\ <0 & \text{if } q < \frac{1}{1+\alpha} \end{cases}$$

Taken together, these asymptotics generate a curve with the shape described.

Proposition 2: When $q > \frac{1}{1+\alpha}$, there exists a unique level of competition μ_2^* at which the gains to exploration, relative to exploitation, are maximized.

Proof: Existence follows from lemma and continuity of the success function. Since the difference of the success function under exploration and exploitation is quadratic in μ, it has at most two real roots, one of which is shown below to be zero, the other of which is shown to be positive if $q \in \left(\frac{1}{1+\alpha}, \frac{1}{2}\right)$ and negative otherwise. Given the shape described by the lemma, the value at which this difference is maximized must be unique.

To find the roots, set $qF\left(\beta_2^H\right) + (1-q)\,F\left(\beta_2^L\right) - F\left(\beta_1\right) = 0$ and solve for μ:

$$
\begin{aligned}
0 &= q\left(\frac{(1+\alpha)\,\beta_1}{(1+\alpha)\,\beta_1+\mu}\right) + (1-q)\left(\frac{\left(1+\frac{1}{\alpha}\right)\beta_1}{\left(1+\frac{1}{\alpha}\right)\beta_1+\mu}\right) - \left(\frac{2\beta_1}{2\beta_1+\mu}\right) \\
&= q\,(1+\alpha)\,\beta_1\left[\left(\left(1+\frac{1}{\alpha}\right)\beta_1+\mu\right)(2\beta_1+\mu)\right] \\
&\quad + (1-q)\left(1+\frac{1}{\alpha}\right)\beta_1\left[\left((1+\alpha)\,\beta_1+\mu\right)(2\beta_1+\mu)\right] \\
&\quad - 2\beta_1\left[\left((1+\alpha)\,\beta_1+\mu\right)\left(\left(1+\frac{1}{\alpha}\right)\beta_1+\mu\right)\right] \\
&= \mu^2\beta_1\left[q\,(1+\alpha)+(1-q)\left(1+\frac{1}{\alpha}\right)-2\right] \\
&\quad + \mu\beta_1^2\left[q\left(2+\alpha+\frac{1}{\alpha}\right)+(1-q)\left(2+\alpha+\frac{1}{\alpha}\right)+2q\,(1+\alpha)+2\,(1-q)\left(1+\frac{1}{\alpha}\right)-2\left(2+\alpha+\frac{1}{\alpha}\right)\right] \\
&\quad + \beta_1^3\left[2q\left(2+\alpha+\frac{1}{\alpha}\right)+2\,(1-q)\left(2+\alpha+\frac{1}{\alpha}\right)-2\left(2+\alpha+\frac{1}{\alpha}\right)\right] \\
&= a\beta_1\mu^2 + b\beta_1^2\mu + c\beta_1^3\,,
\end{aligned}
$$

where

$$
a = q+\alpha q+(1-q)+\frac{1}{\alpha}(1-q)-2 = \alpha q+\frac{1}{\alpha}(1-q)-1 = \frac{1}{\alpha}(\alpha-1)\left((1+\alpha)\,q-1\right)\ \begin{cases} >0 & \text{if } q>\frac{1}{1+\alpha} \\ <0 & \text{if } q<\frac{1}{1+\alpha}\end{cases}
$$

$$
b = 2q+2\alpha q+2\,(1-q)+2\frac{1}{\alpha}(1-q)-\left(2+\alpha+\frac{1}{\alpha}\right) = \frac{1}{\alpha}\left(\alpha^2-1\right)(2q-1)\ \begin{cases} >0 & \text{if } q>\frac{1}{2} \\ <0 & \text{if } q<\frac{1}{2}\end{cases}
$$

$$
c = 0
$$

By the quadratic formula, the roots are thus:

$$\frac{-\left(b\beta_1^2\right) \pm \sqrt{\left(b\beta_1^2\right)^2 - 0}}{2\left(a\beta_1\right)} = \frac{-\left(b\beta_1\right) \pm -\left(b\beta_1\right)}{2a} = \begin{cases} -\beta_1 \frac{b}{a} \\ 0 \end{cases}$$

When $q < \frac{1}{1+\alpha}$, $a < 0$ and $b < 0$, and the non-zero root is negative. When $q \in \left(\frac{1}{1+\alpha}, \frac{1}{2}\right)$, $a > 0$ and $b < 0$, and the non-zero root is positive. When $q > \frac{1}{2}$, $a > 0$ and $b > 0$, and the non-zero root is negative.

Corollary: When $q < \frac{1}{1+\alpha}$, exploration will never be preferred to exploitation.

Proof: Follows from lemma, continuity of the success function, and results from the previous proof showing that when $q < \frac{1}{1+\alpha}$, there is no positive root for the difference of the success function for exploration and exploitation, such that this difference never becomes positive.

Proposition 3: At very low and very high μ, the IR constraint binds: the next-best alternative to exploration is abandonment. At intermediate μ, the IC constraint binds: the next-best alternative is exploitation.

Proof: Lemma 1 can be used to characterize the shape of the gains to exploration versus abandonment and exploitation versus abandonment, since in this model, exploitation is a special case of exploration, with $\alpha = 1$. The proof to Lemma 1 establishes that the gains to exploitation are zero when $\mu = 0$, increasing for small μ, decreasing for large μ, and approach zero from above as $\mu \to \infty$. Provided the prize-normalized cost of exploitation is not greater than the maximum of this function, the payoffs to exploitation will begin negative, turn positive, and finish negative, implying that abandonment (the IR constraint) is binding to exploration at low and high μ and exploitation (the IC constraint) is binding at intermediate μ.

Proposition 4: When $q > \frac{1}{1+\alpha}$, there exists a unique level of competition $\mu^* \in [\mu_1^*, \mu_2^*]$ at which the gains to exploration are maximized relative to the player's next-best alternative.

Proof: Result follows from the first three propositions.

Figure B.1: Exploration, exploitation, and abandonment regions

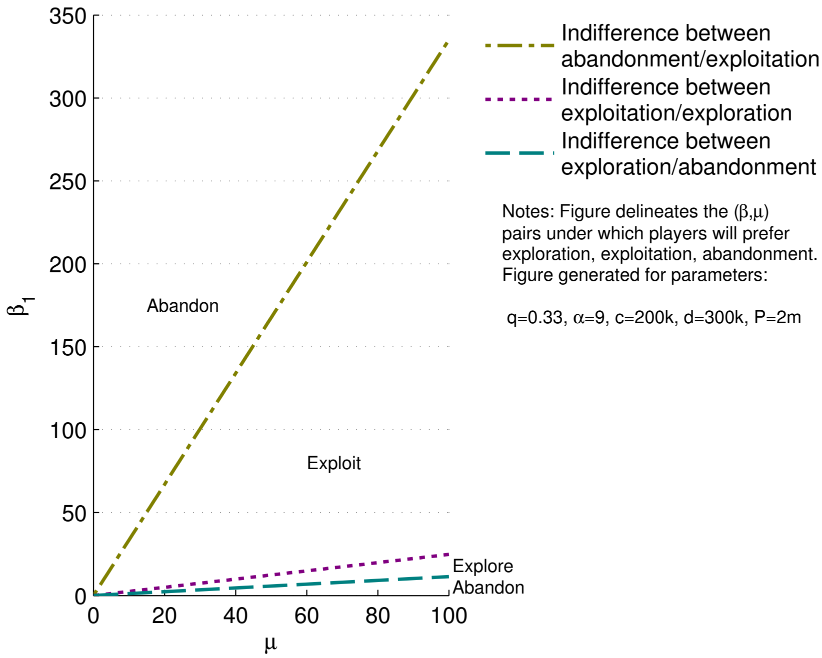


Figure B.2: μ^* at which benefits to exploration are maximized

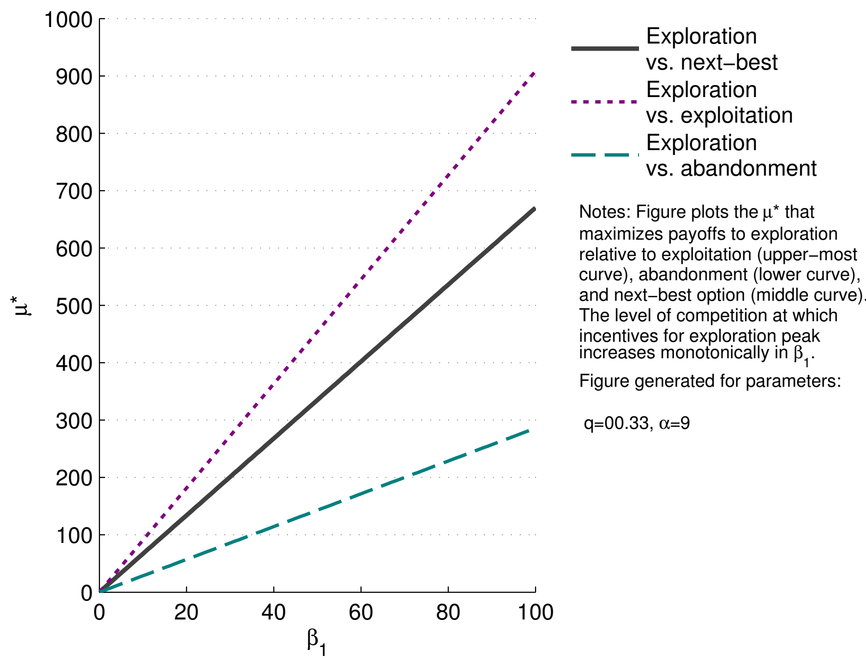
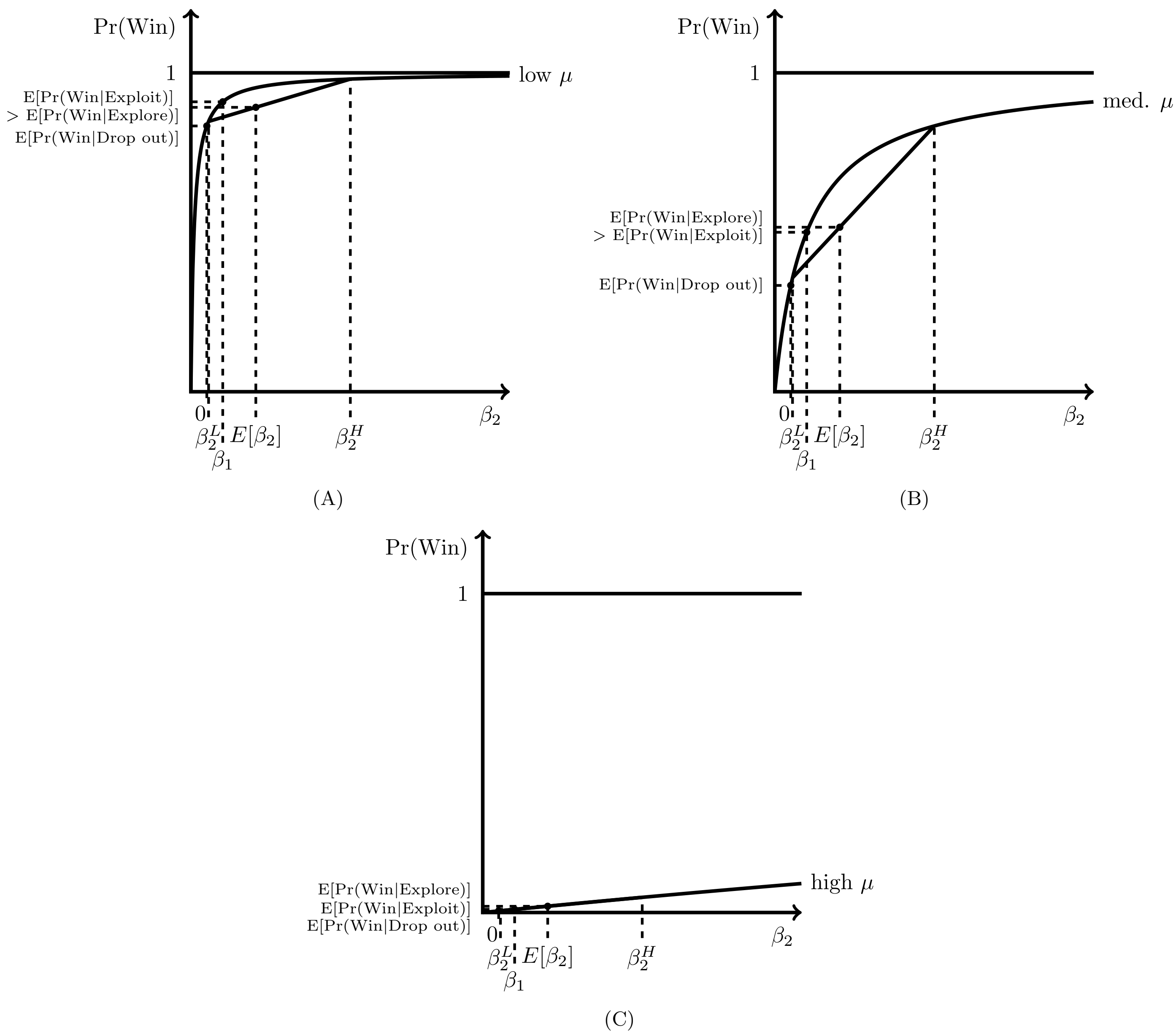


Figure B.3: Gains from experimentation under various states of competition



Notes: Figure illustrates expected gains to exploration, exploitation, and abandonment under low competition (panel A), moderate competition (panel B), and severe competition (panel C). Each subfigure plots a player's probability of winning conditional on the level of competition (μ), the quality of her first design (β_1), and the action taken. The horizontal axis measures the quality of the player's second and final design, which takes one of $\beta_2 \in \{\beta_2^H, \beta_2^L\}$, β_1, or 0. In Panel A, exploitation is preferred to exploration; in panel B, exploration is preferred to exploitation; and in panel C, neither has any significant benefit over abandonment – and any benefit is likely outweighed by the cost. The gains to exploration are determined by the concavity of the success function in the vicinity of β_1, and the patterns illustrated here hold under fairly general conditions described in this section.

Figure generated for $q = 0.33$, $\alpha = 9$, and $\mu = x\beta_1$, for $x = 0.2$, 2, and 200, respectively.

C Dataset Construction

Data were collected on all logo design contests with open (i.e., public) bidding that launched the week of September 3 to 9, 2013, and every three weeks thereafter through the week of November 5 to 11, 2013. Conditional on open bidding, this sample is effectively randomly drawn. The sample used in the paper is further restricted to contests with a single, winner-take-all prize and with no mid-contest rule changes such as prize increases, deadline extensions, and early endings. The sample also excludes one contest that went dormant and resumed after several weeks, as well as a handful of contests whose sponsors simply stopped participating and were never heard from again. These restrictions cause 146 contests to be dropped from the sample. The final dataset includes 122 contests, 4,050 contest-players, and 11,758 designs.

To collect the data, I developed an automated script to scan these contests once an hour for new submissions, save a copy of each design for analysis, and record their owners' identity and performance history from a player profile. I successfully obtained the image files for 96 percent of designs in the final sample. The remaining designs were entered and withdrawn before they could be observed (recall that players can withdraw designs they have entered into a contest, though this option is rarely exercised and can be reversed at the request of a sponsor). All other data were automatically acquired at the conclusion of each contest, once the prize was awarded or the sponsor exercised its outside option of a refund.

C.1 Variables

The dataset includes information on the characteristics of contests, contest-players, and designs:

- Contest-level variables include: the contest sponsor, features of the project brief (title, description, sponsor industry, materials to be included in logo), start and end dates, the prize amount (and whether committed), and the number of players and designs of each rating.

- Contest-player-level variables include: the player's self-reported country, his/her experience in previous contests on the platform (number of contests and designs entered, contests won), and that player's participation and performance in the given contest.

- Design-level variables include: the design's owner, its submission time and order of entry, the feedback it received, the time at which this feedback was given, and whether it was eventually withdrawn. For designs with images acquired, I calculate originality using the procedures described in the next section. The majority of the analysis occurs at the design level.

Note that designs are occasionally re-rated: five percent of all rated designs are re-rated an average of 1.2 times each. Of these, 14 percent are given their original rating, and 83 percent are re-rated within 1 star of the original rating. I treat the first rating on each design to be the most informative, objective measure of quality, since research suggests first instincts tend to be most reliable and ratings revisions are likely made relative to other designs in the contest rather than an objective benchmark.

C.2 Image Comparison Algorithms

This paper uses two distinct algorithms to calculate pairwise similarity scores. One is a perceptual hash algorithm, which creates a digital signature (hash) for each image from its lowest frequency content. As the name implies, a perceptual hash is designed to imitate human perception. The second algorithm is a difference hash, which creates the hash from pixel intensity gradients.

I implement the perceptual hash algorithm and calculate pairwise similarity scores using a variant of the procedure described by the Hacker Factor blog.[1] This requires six steps:

1. Resize each image to 32x32 pixels and convert to grayscale.
2. Compute the discrete cosine transform (DCT) of each image. The DCT is a widely-used transform in signal processing that expresses a finite sequence of data points as a linear combination of cosine functions oscillating at different frequencies. By isolating low frequency content, the DCT reduces a signal (in this case, an image) to its underlying structure. The DCT is broadly used in digital media compression, including MP3 and JPEG formats.
3. Retain the upper-left 16x16 DCT coefficients and calculate the average value, excluding first term.
4. Assign 1s to grid cells with above-average DCT coefficients, and 0s elsewhere.
5. Reshape to 256 bit string; this is the image's digital signature (hash).
6. Compute the Hamming distance between the two hashes and divide by 256.

The similarity score is obtained by subtracting this fraction from one. In a series of sensitivity tests, the perceptual hash algorithm was found to be strongly invariant to transformations in scale, aspect ratio, brightness, and contrast, albeit not rotation. As described, the algorithm will perceive two images that have inverted colors but are otherwise identical to be perfectly dissimilar. I make the algorithm robust to color inversion by comparing each image against the regular and inverted hash of its counterpart in the pair, taking the maximum similarity score, and rescaling so that the scores remain in [0,1]. The resulting score is approximately the absolute value correlation of two images' content.

[1] See http://www.hackerfactor.com/blog/archives/432-Looks-Like-It.html.

I follow a similar procedure outlined by the same blog[2] to implement the difference hash algorithm and calculate an alternative set of similarity scores for robustness checks:

1. Resize each image to 17x16 pixels and convert to grayscale.
2. Calculate horizontal gradient as the change in pixel intensity from left to right, returning a 16x16 grid (note: top to bottom is an equally valid alternative)
3. Assign 1s to grid cells with positive gradient, 0s to cells with negative gradient.
4. Reshape to 256 bit string; this is the image's digital signature (hash).
5. Compute the Hamming distance between the two hashes and divide by 256.

The similarity score is obtained by subtracting this fraction from one. In sensitivity tests, the difference hash algorithm was found to be highly invariant to transformations in scale and aspect ratio, potentially sensitive to changes in brightness and contrast, and very sensitive to rotation. I make the algorithm robust to color inversion using a procedure identical to that described for the perceptual hash.

Though the perceptual and difference hash algorithms are both conceptually and mathematically distinct, and the resulting similarity scores are only modestly correlated ($\rho = 0.38$), the empirical results of Section 3 are qualitatively and quantitatively similar under either algorithm. This consistency is reassurance that the patterns found are not simply an artifact of an arcane image processing algorithm; rather, they appear to be generated by the visual content of the images themselves.

C.3 Why use algorithms?

There are three advantages to using algorithms over human judges. The first is that the algorithms provide a consistent, objective measure of similarity, whereas individuals can have significantly different, subjective perceptions of similarity in practice (Tirilly et al. 2012). This conclusion is supported by a pilot study I attempted using Amazon Mechanical Turk, in which I asked participants to rate the similarity of pairs of images they were shown; the results (not provided here) were generally very noisy, except in cases of nearly identical images, in which case the respondents tended to agree. The second advantage to algorithms over human judges is that algorithms can be directed to evaluate specific features of an image (in this case, the low frequency content or pixel intensity gradient), while human judges will see what they choose to see, and may be attuned to different features in different comparisons. The final advantage of algorithms is more obvious: they are cheap, taking only seconds to execute a comparison.

[2]See http://www.hackerfactor.com/blog/archives/529-Kind-of-Like-That.html.

The evidence of disagreement in subjects' assessments of similarity nevertheless raises a deeper question: is it sensible to apply a uniform similarity measure in this setting? I argue that it is, for the following reasons. First, in both Tirilly et al. (2012) and the Mechanical Turk trials, respondents agreed on extremes, when images were either highly similar or highly dissimilar – in other words, it tends to be obvious when two images are near replicas, which is the margin of variation that matters most for this paper. Squire and Pun (1997) also found that *expert* subjects' assessments of similarity tend to agree at all levels; the designers in this paper could reasonably be classified as visual experts. Finally, divergence in opinion may result from the fact that subjects in the above studies were instructed to assess similarity as they perceive it, rather than in terms of specific features. If subjects were instructed to focus on specific features, they would likely tend to agree – not only with each other, but also with the computer.

C.4 How do the algorithms perform?

In my own experience browsing the designs in the dataset, images that look similar to the naked eye tend to have a high similarity score, particularly under the perceptual hash algorithm. But as Tirilly et al. (2012) show, similarity is in the eye of the beholder – particularly at intermediate levels and when it is being assessed by laypersons. Figure C.1 illustrates the performance of the algorithms for three logos entered in the order shown by the same player in one contest (not necessarily from the sampled platform):

Figure C.1: Performance of image comparison algorithms

(1) (2) (3)

Notes: 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 (2) and (3). The difference hash algorithm calculates similarity scores of 0.508 for (1) and (2) and 0.891 for (2) and (3).

The first two images have several features in common but also have some notable differences. Each is centered, defined by a circular frame with text underneath, and presented against a similar backdrop. However the content of the circular frame and the font of the text below are considerably different, and the first logo is in black and white while the second one is in color. The perceptual hash algorithm assigns these two logos a similarity score of 31 percent, while the difference hash gives them 51 percent.

In contrast, the second two images appear much more similar. They again have similar layouts, but now they share the same color assignments and the same content in the frame. Lesser differences remain, primarily with respect to the font style, but the logos appear broadly similar. The perceptual hash algorithm assigns these two logos a similarity score of 71 percent; the difference hash, 89 percent.

The algorithms thus pass the gut check in this example, which is not particularly unique: further examples using better-known brands are provided below. In light of this evidence, and the consistency of the paper's results, I believe that these algorithms provide empirically valid measures of experimentation.

Figure C.2: Volkswagen logo in 1939, 1967, 1995, 1999

(1939) (1967) (1995) (1999)

Notes: Figure shows the evolution of Volkswagen logos since 1939. The perceptual hash algorithm calculates similarity scores of 0.055 for the 1937 and 1967 logos, 0.430 for the 1967 and 1995 logos, and 0.844 for the 1995 and 1999 logos. The difference hash algorithm calculates similarity scores of 0.195, 0.539, and 0.953, respectively.

Figure C.3: Microsoft Windows 95, XP, 7, and 8 logos

(Windows 95) (Windows XP) (Windows 7) (Windows 8)

Notes: Figure shows a sequence of Windows logos. The perceptual hash algorithm calculates similarity scores of 0.195 for the Windows 95 and XP logos, 0.531 for the Windows XP and 7 logos, and 0.148 for the Windows 7 and 8 logos. The difference hash algorithm calculates similarity scores of 0.055, 0.563, and 0.117, respectively. The reason why the similarity of the Windows XP and 7 logos is not evaluated to be even higher is because the contrast generated by the latter's spotlight and shadow changes the structure of the image (for example, it changes the intensity gradient calculated by the difference hash algorithm).

Appendix References:

[1] Tirilly, Pierre, Chunsheng Huang, Wooseob Jeong, Xiangming Mu, Iris Xie, and Jin Zhang. 2012. "Image Similarity as Assessed by Users: A Quantitative Study." *Proceedings of the American Society for Information Science and Technology*, 49(1), pp. 1-10.

[2] Squire, David and Thierry Pun. 1997. "A Comparison of Human and Machine Assessments of Image Similarity for the Organization of Image Databases." *Proceedings of the Scandinavian Conference on Image Analysis*, Lappeenranta, Finland.

D Robustness Checks (1)

The following tables provide variants of the tables in Section 3 estimating the effects of feedback and competition on experimentation, using the difference hash algorithm instead of the preferred, perceptual hash algorithm. These estimates serve as robustness checks to the principal empirical results of the paper, demonstrating that they are not sensitive to the procedure used to calculate similarity scores.

Table D.1 is a robustness check on Table 7; Table D.2, on Table 8; Table D.3, on Table 9; Table D.4, on Table 10; and Table D.5, on Table 11. The results in these appendix tables are qualitatively and quantitatively similar to those in the body of the paper.

Table D.1: Similarity to player's previous designs (difference hash)

	(1)	(2)	(3)	(4)
Player's prior best rating==5	0.270***	0.253***	0.269***	0.256***
	(0.086)	(0.087)	(0.087)	(0.086)
* 1+ competing 5-stars	-0.128**	-0.141**	-0.127**	-0.140**
	(0.058)	(0.056)	(0.058)	(0.056)
* prize value ($100s)	-0.038	-0.047*	-0.034	-0.046*
	(0.026)	(0.027)	(0.026)	(0.027)
Player's prior best rating==4	0.057***	0.025	0.064***	0.030
	(0.020)	(0.021)	(0.020)	(0.021)
Player's prior best rating==3	0.027*	0.011	0.035**	0.016
	(0.017)	(0.017)	(0.017)	(0.017)
Player's prior best rating==2	-0.004	-0.012	0.003	-0.008
	(0.020)	(0.020)	(0.020)	(0.020)
One or more competing 5-stars	-0.011	-0.022	-0.011	-0.022
	(0.022)	(0.024)	(0.022)	(0.023)
Days remaining	-0.004	0.001	-0.004*	0.001
	(0.003)	(0.007)	(0.003)	(0.007)
Constant	0.508***	0.454***	0.512***	0.461***
	(0.139)	(0.160)	(0.139)	(0.159)
N	5075	5075	5075	5075
R^2	0.48	0.48	0.48	0.48
Controls	No	Yes	No	Yes
Contest FEs	Yes	Yes	Yes	Yes
Player FEs	Yes	Yes	Yes	Yes
Forthcoming ratings	No	No	Yes	Yes

Notes: Table shows the effects of feedback on players' experimentation. Observations are designs. Dependent variable is a continuous measure of a design's maximum similarity to previous entries in the same contest by the same player, taking values in [0,1], where a value of 1 indicates the design is identical to another. The mean value of this variable in the sample is 0.58 (s.d. 0.28). Columns (2) and (4) control for time of submission and number of previous designs entered by the player and her competitors. Columns (3) and (4) additionally control for the best *forthcoming* rating on the player's not-yet-rated designs. Similarity scores in this table are calculated using a difference hash algorithm. Preceding designs/ratings are defined to be those entered/provided at least 60 minutes prior to the given design. *, **, *** represent significance at the 0.1, 0.05, and 0.01 levels, respectively. SEs clustered by player in parentheses.

Table D.2: Similarity to player's best previously-rated designs & intra-batch similarity (difference hash)

	Designs		Batches (uwtd.)		Batches (wtd.)	
	(1)	(2)	(3)	(4)	(5)	(6)
Player's prior best rating==5	0.244*	0.242*	0.224	0.246	0.236	0.260
	(0.131)	(0.141)	(0.299)	(0.293)	(0.286)	(0.281)
* 1+ competing 5-stars	-0.168*	-0.177**	-0.327**	-0.324**	-0.314**	-0.308**
	(0.086)	(0.087)	(0.146)	(0.144)	(0.147)	(0.145)
* prize value ($100s)	-0.018	-0.024	-0.022	-0.023	-0.025	-0.026
	(0.038)	(0.042)	(0.093)	(0.092)	(0.087)	(0.085)
Player's prior best rating==4	0.066*	0.049	-0.016	-0.003	-0.012	0.004
	(0.039)	(0.041)	(0.031)	(0.032)	(0.029)	(0.031)
Player's prior best rating==3	0.044	0.033	0.011	0.019	0.010	0.020
	(0.038)	(0.039)	(0.033)	(0.035)	(0.031)	(0.032)
Player's prior best rating==2	0.014	0.007	-0.019	-0.012	-0.021	-0.014
	(0.040)	(0.040)	(0.047)	(0.049)	(0.045)	(0.045)
One or more competing 5-stars	-0.012	-0.019	-0.018	-0.018	-0.014	-0.017
	(0.031)	(0.032)	(0.033)	(0.034)	(0.031)	(0.033)
Days remaining	0.002	0.001	-0.000	-0.002	-0.000	0.001
	(0.003)	(0.009)	(0.004)	(0.008)	(0.004)	(0.009)
Constant	0.844***	0.863***	0.646***	0.673***	0.672***	0.661***
	(0.152)	(0.173)	(0.121)	(0.156)	(0.096)	(0.128)
N	3871	3871	1987	1987	1987	1987
R^2	0.53	0.53	0.59	0.59	0.59	0.59
Controls	No	Yes	No	Yes	No	Yes
Contest FEs	Yes	Yes	Yes	Yes	Yes	Yes
Player FEs	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Table shows the effects of feedback on players'experimentation. Observations in Columns (1) and (2) are designs, and 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. The mean value of this variable in the sample is 0.52 (s.d. 0.30). Observations in Columns (3) to (6) are design batches, which are defined to be a set of designs by a single player entered into a contest in close proximity (15 minutes), and dependent variable is a continuous measure of intra-batch similarity, taking values in [0,1], where a value of 1 indicates that two designs in the batch are identical. The mean value of this variable in the sample is 0.72 (s.d. 0.27). Columns (5) and (6) weight the batch regressions by batch size. All columns control for the time of submission and number of previous designs entered by the player and her competitors. Similarity scores in this table are calculated using a difference hash algorithm. Preceding designs/ratings are defined to be those entered/provided at least 60 minutes prior to the given design. *, **, *** represent significance at the 0.1, 0.05, and 0.01 levels, respectively. SEs clustered by player in parentheses.

Table D.3: Change in similarity to player's best previously-rated designs (difference hash)

	(1)	(2)	(3)	(4)	(5)	(6)
Δ(Player's best rating==5)	0.657***	0.680***	0.687**	0.659***	0.693***	0.693***
	(0.218)	(0.256)	(0.269)	(0.217)	(0.256)	(0.267)
* 1+ competing 5-stars	-0.347**	-0.374*	-0.362*	-0.350**	-0.379*	-0.368*
	(0.174)	(0.206)	(0.218)	(0.174)	(0.206)	(0.218)
* prize value ($100s)	-0.049	-0.060	-0.063	-0.048	-0.063	-0.064
	(0.046)	(0.055)	(0.057)	(0.046)	(0.055)	(0.057)
Δ(Player's best rating==4)	0.262***	0.236***	0.231***	0.262***	0.237***	0.232***
	(0.070)	(0.081)	(0.086)	(0.070)	(0.081)	(0.086)
Δ(Player's best rating==3)	0.192***	0.169**	0.161**	0.192***	0.169**	0.162**
	(0.062)	(0.073)	(0.077)	(0.063)	(0.073)	(0.077)
Δ(Player's best rating==2)	0.132**	0.110	0.104	0.131**	0.110	0.104
	(0.058)	(0.067)	(0.071)	(0.058)	(0.067)	(0.071)
One or more competing 5-stars	-0.005	-0.000	-0.005	0.001	0.000	-0.001
	(0.016)	(0.016)	(0.025)	(0.018)	(0.016)	(0.029)
Days remaining	-0.000	-0.000	0.000	-0.007	-0.005	-0.009
	(0.002)	(0.002)	(0.003)	(0.005)	(0.004)	(0.009)
Constant	-0.012	-0.013	-0.237***	0.058	0.038	-0.169**
	(0.010)	(0.010)	(0.045)	(0.050)	(0.047)	(0.072)
N	2694	2694	2694	2694	2694	2694
R^2	0.04	0.10	0.13	0.04	0.10	0.13
Controls	No	No	No	Yes	Yes	Yes
Contest FEs	Yes	No	Yes	Yes	No	Yes
Player FEs	No	Yes	Yes	No	Yes	Yes

Notes: Table shows the effects of feedback on players' experimentation. Observations are designs. 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 experimenting to copying (and a value of -1, the converse). The mean value of this variable in the sample is -0.01 (s.d. 0.25). Columns (4) to (6) control for time of submission and number of previous designs entered by the player and competitors. Similarity scores in this table are calculated using a difference hash algorithm. Preceding designs/ratings are defined to be those entered/provided at least 60 minutes prior to the given design. *, **, *** represent significance at the 0.1, 0.05, and 0.01 levels, respectively. SEs clustered by player in parentheses.

Table D.4: Similarity to player's best not-yet-rated designs (placebo test; using difference hash)

	Similarity to forthcoming			Residual
	(1)	(2)	(3)	(4)
Player's best forthcoming rating==5	0.512	0.215	0.122	0.238
	(0.447)	(0.214)	(0.202)	(0.207)
* 1+ competing 5-stars	-0.220	-0.111	-0.052	-0.108
	(0.262)	(0.144)	(0.146)	(0.145)
* prize value ($100s)	-0.122	-0.026	-0.006	-0.024
	(0.089)	(0.049)	(0.046)	(0.051)
Player's best forthcoming rating==4	0.105	0.105	0.109	0.122
	(0.085)	(0.113)	(0.117)	(0.117)
Player's best forthcoming rating==3	0.093	0.103	0.101	0.121
	(0.057)	(0.135)	(0.140)	(0.143)
Player's best forthcoming rating==2	0.045	0.049	0.048	0.077
	(0.050)	(0.136)	(0.142)	(0.139)
One or more competing 5-stars	-0.076	-0.079	-0.081	-0.075
	(0.081)	(0.122)	(0.126)	(0.131)
Days remaining	-0.019	0.003	0.006	0.004
	(0.030)	(0.055)	(0.056)	(0.062)
Constant	0.998***	0.452	0.348	-0.179
	(0.170)	(0.473)	(0.513)	(0.489)
N	1147	577	577	577
R^2	0.69	0.87	0.88	0.69
Controls	Yes	Yes	Yes	Yes
Contest FEs	Yes	Yes	Yes	Yes
Player FEs	Yes	Yes	Yes	Yes

Notes: Table provides a test of the effects of not-yet-available feedback on players' experimentation. Observations are designs. Dependent variable in Columns (1) to (3) is a continuous measure of a design's similarity to the best designs that the player has previously entered and 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. The mean value of this variable in the sample is 0.50 (s.d. 0.29). If players depend on sponsors' ratings for signals of quality, then forthcoming ratings should have no effect on current experimentation. The results of Column (1) suggest this may not be the case; however, similarity to an unrated design may actually be the result of both these designs being tweaks on a third design. To account for this possibility, Column (2) controls for the given design's similarity to the best previously-rated design, the best not-yet-rated design's similarity to the best previously-rated design, and their interaction. Column (3) allows these controls to vary by the best rating previously received. 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 placebo that is not explained by jointly-occurring imitation of a third design. All columns control for time of submission and number of previous designs entered by the player and her competitors. Similarity scores in this table are calculated using a difference hash algorithm. Preceding designs/ratings are defined to be those entered/provided at least 60 minutes prior to the given design. *, **, *** represent significance at the 0.1, 0.05, and 0.01 levels, respectively. SEs clustered by player in parentheses.

Table D.5: Similarity and change in similarity to competitors' best previously-rated designs (difference hash)

	(1)	(2)		(3)	(4)
Competitors' best==5	-0.168***	-0.127***	Δ(Competitors' best==5)	-0.064	-0.062
	-0.033	-0.034		-0.096	(0.096)
* 1+ own 5-stars	-0.018	-0.017	* 1+ own 5-stars	0.006	0.005
	-0.017	-0.017		-0.066	(0.066)
* prize value ($100s)	0.025***	0.005	* prize value ($100s)	0.001	-0.000
	-0.006	-0.006		-0.018	(0.018)
Competitors' best==4	0.027	0.033	Δ(Competitors' best==4)	0.069	0.069
	-0.025	-0.025		-0.076	(0.076)
Competitors' best==3	0.032	0.045*	Δ(Competitors' best==3)	0.068	0.069
	-0.025	-0.025		-0.075	(0.076)
Competitors' best==2	-0.052*	-0.052*	Δ(Competitors' best==2)	0.014	0.016
	-0.03	-0.029		-0.077	(0.077)
One or more own 5-stars	0.004	0.005	One or more own 5-stars	0.016	0.020
	-0.027	-0.03		-0.012	(0.013)
Days remaining	-0.011***	0.008***	Days remaining	0.001	0.004
	-0.001	-0.002		-0.001	(0.003)
Constant	0.537***	0.396***	Constant	0.035	0.016
	-0.094	-0.097		-0.114	(0.118)
N	9709	9709	N	6065	6065
R^2	0.54	0.54	R^2	0.14	0.15
Controls	No	Yes	Controls	No	Yes
Contest FEs	Yes	Yes	Contest FEs	Yes	Yes
Player FEs	Yes	Yes	Player FEs	Yes	Yes

Notes: Table provides a test of players' ability to discern the quality of, and then imitate, competing designs. Observations are designs. Dependent variable in Columns (1) and (2) is a continuous measure of the design's similarity to the highest-rated preceding entries by *other* players, taking values in [0,1], where a value of 1 indicates that the design is identical to another. The mean value in the sample is 0.33 (s.d. 0.21). Dependent variable in Columns (3) and (4) is a continuous measure of the *change* in designs' similarity to the highest-rated preceding entries by other players, taking values in [-1,1], where a value of 0 indicates that the player's current design is equally similar to the best competing design as was her previous design, and a value of 1 indicates that the player transitioned fully from experimenting to copying (and a value of -1, the converse). The mean value of this variable in the sample is 0.00 (s.d. 0.15). In general, players are provided only the distribution of ratings on competing designs; ratings of specific competing designs are not observed. Results in this table test whether players can nevertheless identify and imitate leading competition. Columns (2) and (4) control for time of submission and number of previous designs entered by the player and her competitors. Similarity scores in this table are calculated using a difference hash algorithm. Preceding designs/ratings are defined to be those entered/provided at least 60 minutes prior to the given design. *, **, *** represent significance at the 0.1, 0.05, and 0.01 levels, respectively. Robust SEs in parentheses.

E Robustness Checks (2)

In additional robustness checks, I show that competition has a constant effect on high-performing players' tendency to experiment. Tables E.1 to E.3 demonstrate this result with the perceptual hash similarity measures, and Tables E.4 to E.6 do so with the difference hash measures. In all cases, I estimate differential effects for one vs. multiple top-rated, competing designs and find no differential effect.

Table E.1: Similarity to player's previous designs (p. hash)

	(1)	(2)	(3)	(4)
Player's prior best rating==5	0.269***	0.250***	0.268***	0.253***
	(0.087)	(0.087)	(0.088)	(0.086)
* 1+ competing 5-stars	-0.102	-0.110	-0.101	-0.108
	(0.087)	(0.080)	(0.087)	(0.081)
* 2+ competing 5-stars	-0.034	-0.039	-0.034	-0.041
	(0.094)	(0.080)	(0.093)	(0.081)
* prize value ($100s)	-0.037	-0.046*	-0.034	-0.045*
	(0.026)	(0.027)	(0.027)	(0.027)
Player's prior best rating==4	0.058***	0.027	0.066***	0.032
	(0.020)	(0.021)	(0.020)	(0.021)
Player's prior best rating==3	0.028*	0.013	0.036**	0.018
	(0.017)	(0.017)	(0.017)	(0.017)
Player's prior best rating==2	-0.004	-0.012	0.003	-0.008
	(0.020)	(0.020)	(0.020)	(0.020)
One or more competing 5-stars	-0.042	-0.044	-0.043	-0.045
	(0.036)	(0.038)	(0.036)	(0.037)
Two or more competing 5-stars	0.049	0.036	0.050	0.037
	(0.037)	(0.040)	(0.038)	(0.040)
Days remaining	-0.004	0.001	-0.004	0.001
	(0.003)	(0.007)	(0.002)	(0.007)
Constant	0.506***	0.460***	0.510***	0.467***
	(0.139)	(0.160)	(0.138)	(0.159)
N	5075	5075	5075	5075
R^2	0.48	0.48	0.48	0.48
Controls	No	Yes	No	Yes
Contest FEs	Yes	Yes	Yes	Yes
Player FEs	Yes	Yes	Yes	Yes
Forthcoming ratings	No	No	Yes	Yes

Notes: Table shows the effects of feedback on players' experimentation. Observations are designs. Dependent variable is a continuous measure of a design's maximum similarity to previous entries in the same contest by the same player, taking values in [0,1], where a value of 1 indicates the design is identical to another. The mean value of this variable in the sample is 0.58 (s.d. 0.28). Columns (2) and (4) control for time of submission and number of previous designs entered by the player and her competitors. Columns (3) and (4) additionally control for the best *forthcoming* rating on the player's not-yet-rated designs. Similarity scores in this table are calculated using a difference hash algorithm. Preceding designs/ratings are defined to be those entered/provided at least 60 minutes prior to the given design. *, **, *** represent significance at the 0.1, 0.05, and 0.01 levels, respectively. SEs clustered by player.

Table E.2: Similarity to player's best previously-rated designs & intra-batch similarity (p. hash)

	Designs		Batches (uwtd.)		Batches (wtd.)	
	(1)	(2)	(3)	(4)	(5)	(6)
Player's prior best rating==5	0.239*	0.234*	0.239	0.271	0.253	0.285
	(0.132)	(0.141)	(0.317)	(0.312)	(0.300)	(0.295)
* 1+ competing 5-stars	-0.144	-0.149	-0.327**	-0.328**	-0.328**	-0.327**
	(0.140)	(0.138)	(0.143)	(0.142)	(0.131)	(0.131)
* 2+ competing 5-stars	-0.032	-0.036	0.006	0.009	0.025	0.030
	(0.143)	(0.134)	(0.189)	(0.186)	(0.173)	(0.172)
* prize value ($100s)	-0.016	-0.022	-0.026	-0.031	-0.030	-0.033
	(0.038)	(0.042)	(0.100)	(0.099)	(0.092)	(0.090)
Player's prior best rating==4	0.067*	0.050	-0.015	-0.002	-0.012	0.004
	(0.039)	(0.041)	(0.031)	(0.032)	(0.029)	(0.031)
Player's prior best rating==3	0.044	0.033	0.012	0.019	0.011	0.020
	(0.038)	(0.039)	(0.033)	(0.035)	(0.031)	(0.032)
Player's prior best rating==2	0.013	0.006	-0.018	-0.011	-0.020	-0.012
	(0.040)	(0.040)	(0.047)	(0.048)	(0.044)	(0.045)
One or more competing 5-stars	-0.050	-0.050	0.037	0.038	0.031	0.030
	(0.045)	(0.045)	(0.041)	(0.042)	(0.037)	(0.038)
Two or more competing 5-stars	0.058	0.052	-0.089*	-0.096*	-0.076	-0.082
	(0.049)	(0.051)	(0.052)	(0.055)	(0.048)	(0.051)
Days remaining	0.002	0.001	-0.001	-0.003	-0.001	-0.000
	(0.003)	(0.009)	(0.004)	(0.009)	(0.004)	(0.009)
Constant	0.841***	0.870***	0.649***	0.677***	0.675***	0.666***
	(0.152)	(0.174)	(0.120)	(0.156)	(0.095)	(0.128)
N	3871	3871	1987	1987	1987	1987
R^2	0.53	0.54	0.59	0.59	0.59	0.59
Controls	No	Yes	No	Yes	No	Yes
Contest FEs	Yes	Yes	Yes	Yes	Yes	Yes
Player FEs	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Table shows the effects of feedback on players'experimentation. Observations in Columns (1) and (2) are designs, and 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. The mean value of this variable in the sample is 0.52 (s.d. 0.30). Observations in Columns (3) to (6) are design batches, which are defined to be a set of designs by a single player entered into a contest in close proximity (15 minutes), and dependent variable is a continuous measure of intra-batch similarity, taking values in [0,1], where a value of 1 indicates that two designs in the batch are identical. The mean value of this variable in the sample is 0.72 (s.d. 0.27). Columns (5) and (6) weight the batch regressions by batch size. All columns control for the time of submission and number of previous designs entered by the player and her competitors. Similarity scores in this table are calculated using a difference hash algorithm. Preceding designs/ratings are defined to be those entered/provided at least 60 minutes prior to the given design. *, **, *** represent significance at the 0.1, 0.05, and 0.01 levels, respectively. SEs clustered by player in parentheses.

Table E.3: Change in similarity to player's best previously-rated designs (p. hash)

	(1)	(2)	(3)	(4)	(5)	(6)
Δ(Player's best rating==5)	0.656***	0.685***	0.693***	0.657***	0.695***	0.695***
	(0.218)	(0.253)	(0.265)	(0.217)	(0.253)	(0.264)
* 1+ competing 5-stars	-0.398*	-0.449*	-0.425*	-0.400*	-0.456*	-0.429*
	(0.210)	(0.235)	(0.255)	(0.211)	(0.236)	(0.256)
* 2+ competing 5-stars	0.070	0.106	0.090	0.068	0.109	0.088
	(0.154)	(0.167)	(0.185)	(0.155)	(0.167)	(0.185)
* prize value ($100s)	-0.048	-0.060	-0.064	-0.047	-0.062	-0.063
	(0.046)	(0.053)	(0.055)	(0.045)	(0.053)	(0.055)
Δ(Player's best rating==4)	0.264***	0.240***	0.235***	0.264***	0.241***	0.237***
	(0.070)	(0.081)	(0.086)	(0.070)	(0.081)	(0.085)
Δ(Player's best rating==3)	0.194***	0.174**	0.165**	0.194***	0.173**	0.166**
	(0.063)	(0.073)	(0.077)	(0.063)	(0.074)	(0.077)
Δ(Player's best rating==2)	0.134**	0.114*	0.108	0.133**	0.114*	0.108
	(0.058)	(0.068)	(0.071)	(0.058)	(0.068)	(0.071)
One or more competing 5-stars	-0.013	-0.020	-0.031	-0.008	-0.018	-0.026
	(0.029)	(0.035)	(0.047)	(0.029)	(0.036)	(0.048)
Two or more competing 5-stars	0.011	0.028	0.041	0.015	0.026	0.046
	(0.027)	(0.037)	(0.047)	(0.028)	(0.038)	(0.048)
Days remaining	-0.000	-0.000	0.001	-0.007	-0.005	-0.009
	(0.002)	(0.002)	(0.003)	(0.005)	(0.004)	(0.009)
Constant	-0.013	-0.014	-0.243***	0.059	0.040	-0.171**
	(0.010)	(0.010)	(0.044)	(0.049)	(0.046)	(0.073)
N	2694	2694	2694	2694	2694	2694
R^2	0.04	0.10	0.13	0.04	0.10	0.13
Controls	No	No	No	Yes	Yes	Yes
Contest FEs	Yes	No	Yes	Yes	No	Yes
Player FEs	No	Yes	Yes	No	Yes	Yes

Notes: Table shows the effects of feedback on players' experimentation. Observations are designs. 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 experimenting to copying (and a value of -1, the converse). The mean value of this variable in the sample is -0.01 (s.d. 0.25). Columns (4) to (6) control for time of submission and number of previous designs entered by the player and competitors. Similarity scores in this table are calculated using a difference hash algorithm. Preceding designs/ratings are defined to be those entered/provided at least 60 minutes prior to the given design. *, **, *** represent significance at the 0.1, 0.05, and 0.01 levels, respectively. SEs clustered by player in parentheses.

Table E.4: Similarity to player's previous designs (d. hash)

	(1)	(2)	(3)	(4)
Player's prior best rating==5	0.269***	0.250***	0.268***	0.253***
	(0.087)	(0.087)	(0.088)	(0.086)
* 1+ competing 5-stars	-0.102	-0.110	-0.101	-0.108
	(0.087)	(0.080)	(0.087)	(0.081)
* 2+ competing 5-stars	-0.034	-0.039	-0.034	-0.041
	(0.094)	(0.080)	(0.093)	(0.081)
* prize value ($100s)	-0.037	-0.046*	-0.034	-0.045*
	(0.026)	(0.027)	(0.027)	(0.027)
Player's prior best rating==4	0.058***	0.027	0.066***	0.032
	(0.020)	(0.021)	(0.020)	(0.021)
Player's prior best rating==3	0.028*	0.013	0.036**	0.018
	(0.017)	(0.017)	(0.017)	(0.017)
Player's prior best rating==2	-0.004	-0.012	0.003	-0.008
	(0.020)	(0.020)	(0.020)	(0.020)
One or more competing 5-stars	-0.042	-0.044	-0.043	-0.045
	(0.036)	(0.038)	(0.036)	(0.037)
Two or more competing 5-stars	0.049	0.036	0.050	0.037
	(0.037)	(0.040)	(0.038)	(0.040)
Days remaining	-0.004	0.001	-0.004	0.001
	(0.003)	(0.007)	(0.002)	(0.007)
Constant	0.506***	0.460***	0.510***	0.467***
	(0.139)	(0.160)	(0.138)	(0.159)
N	5075	5075	5075	5075
R^2	0.48	0.48	0.48	0.48
Controls	No	Yes	No	Yes
Contest FEs	Yes	Yes	Yes	Yes
Player FEs	Yes	Yes	Yes	Yes
Forthcoming ratings	No	No	Yes	Yes

Notes: Table shows the effects of feedback on players' experimentation. Observations are designs. Dependent variable is a continuous measure of a design's maximum similarity to previous entries in the same contest by the same player, taking values in [0,1], where a value of 1 indicates the design is identical to another. The mean value of this variable in the sample is 0.58 (s.d. 0.28). Columns (2) and (4) control for time of submission and number of previous designs entered by the player and her competitors. Columns (3) and (4) additionally control for the best *forthcoming* rating on the player's not-yet-rated designs. Similarity scores in this table are calculated using a difference hash algorithm. Preceding designs/ratings are defined to be those entered/provided at least 60 minutes prior to the given design. *, **, *** represent significance at the 0.1, 0.05, and 0.01 levels, respectively. SEs clustered by player.

Table E.5: Similarity to player's best previously-rated designs & intra-batch similarity (d. hash)

	Designs		Batches (uwtd.)		Batches (wtd.)	
	(1)	(2)	(3)	(4)	(5)	(6)
Player's prior best rating==5	0.239*	0.234*	0.239	0.271	0.253	0.285
	(0.132)	(0.141)	(0.317)	(0.312)	(0.300)	(0.295)
* 1+ competing 5-stars	-0.144	-0.149	-0.327**	-0.328**	-0.328**	-0.327**
	(0.140)	(0.138)	(0.143)	(0.142)	(0.131)	(0.131)
* 2+ competing 5-stars	-0.032	-0.036	0.006	0.009	0.025	0.030
	(0.143)	(0.134)	(0.189)	(0.186)	(0.173)	(0.172)
* prize value ($100s)	-0.016	-0.022	-0.026	-0.031	-0.030	-0.033
	(0.038)	(0.042)	(0.100)	(0.099)	(0.092)	(0.090)
Player's prior best rating==4	0.067*	0.050	-0.015	-0.002	-0.012	0.004
	(0.039)	(0.041)	(0.031)	(0.032)	(0.029)	(0.031)
Player's prior best rating==3	0.044	0.033	0.012	0.019	0.011	0.020
	(0.038)	(0.039)	(0.033)	(0.035)	(0.031)	(0.032)
Player's prior best rating==2	0.013	0.006	-0.018	-0.011	-0.020	-0.012
	(0.040)	(0.040)	(0.047)	(0.048)	(0.044)	(0.045)
One or more competing 5-stars	-0.050	-0.050	0.037	0.038	0.031	0.030
	(0.045)	(0.045)	(0.041)	(0.042)	(0.037)	(0.038)
Two or more competing 5-stars	0.058	0.052	-0.089*	-0.096*	-0.076	-0.082
	(0.049)	(0.051)	(0.052)	(0.055)	(0.048)	(0.051)
Days remaining	0.002	0.001	-0.001	-0.003	-0.001	-0.000
	(0.003)	(0.009)	(0.004)	(0.009)	(0.004)	(0.009)
Constant	0.841***	0.870***	0.649***	0.677***	0.675***	0.666***
	(0.152)	(0.174)	(0.120)	(0.156)	(0.095)	(0.128)
N	3871	3871	1987	1987	1987	1987
R^2	0.53	0.54	0.59	0.59	0.59	0.59
Controls	No	Yes	No	Yes	No	Yes
Contest FEs	Yes	Yes	Yes	Yes	Yes	Yes
Player FEs	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Table shows the effects of feedback on players'experimentation. Observations in Columns (1) and (2) are designs, and 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. The mean value of this variable in the sample is 0.52 (s.d. 0.30). Observations in Columns (3) to (6) are design batches, which are defined to be a set of designs by a single player entered into a contest in close proximity (15 minutes), and dependent variable is a continuous measure of intra-batch similarity, taking values in [0,1], where a value of 1 indicates that two designs in the batch are identical. The mean value of this variable in the sample is 0.72 (s.d. 0.27). Columns (5) and (6) weight the batch regressions by batch size. All columns control for the time of submission and number of previous designs entered by the player and her competitors. Similarity scores in this table are calculated using a difference hash algorithm. Preceding designs/ratings are defined to be those entered/provided at least 60 minutes prior to the given design. *, **, *** represent significance at the 0.1, 0.05, and 0.01 levels, respectively. SEs clustered by player in parentheses.

Table E.6: Change in similarity to player's best previously-rated designs (d. hash)

	(1)	(2)	(3)	(4)	(5)	(6)
Δ(Player's best rating==5)	0.656***	0.685***	0.693***	0.657***	0.695***	0.695***
	(0.218)	(0.253)	(0.265)	(0.217)	(0.253)	(0.264)
* 1+ competing 5-stars	-0.398*	-0.449*	-0.425*	-0.400*	-0.456*	-0.429*
	(0.210)	(0.235)	(0.255)	(0.211)	(0.236)	(0.256)
* 2+ competing 5-stars	0.070	0.106	0.090	0.068	0.109	0.088
	(0.154)	(0.167)	(0.185)	(0.155)	(0.167)	(0.185)
* prize value ($100s)	-0.048	-0.060	-0.064	-0.047	-0.062	-0.063
	(0.046)	(0.053)	(0.055)	(0.045)	(0.053)	(0.055)
Δ(Player's best rating==4)	0.264***	0.240***	0.235***	0.264***	0.241***	0.237***
	(0.070)	(0.081)	(0.086)	(0.070)	(0.081)	(0.085)
Δ(Player's best rating==3)	0.194***	0.174**	0.165**	0.194***	0.173**	0.166**
	(0.063)	(0.073)	(0.077)	(0.063)	(0.074)	(0.077)
Δ(Player's best rating==2)	0.134**	0.114*	0.108	0.133**	0.114*	0.108
	(0.058)	(0.068)	(0.071)	(0.058)	(0.068)	(0.071)
One or more competing 5-stars	-0.013	-0.020	-0.031	-0.008	-0.018	-0.026
	(0.029)	(0.035)	(0.047)	(0.029)	(0.036)	(0.048)
Two or more competing 5-stars	0.011	0.028	0.041	0.015	0.026	0.046
	(0.027)	(0.037)	(0.047)	(0.028)	(0.038)	(0.048)
Days remaining	-0.000	-0.000	0.001	-0.007	-0.005	-0.009
	(0.002)	(0.002)	(0.003)	(0.005)	(0.004)	(0.009)
Constant	-0.013	-0.014	-0.243***	0.059	0.040	-0.171**
	(0.010)	(0.010)	(0.044)	(0.049)	(0.046)	(0.073)
N	2694	2694	2694	2694	2694	2694
R^2	0.04	0.10	0.13	0.04	0.10	0.13
Controls	No	No	No	Yes	Yes	Yes
Contest FEs	Yes	No	Yes	Yes	No	Yes
Player FEs	No	Yes	Yes	No	Yes	Yes

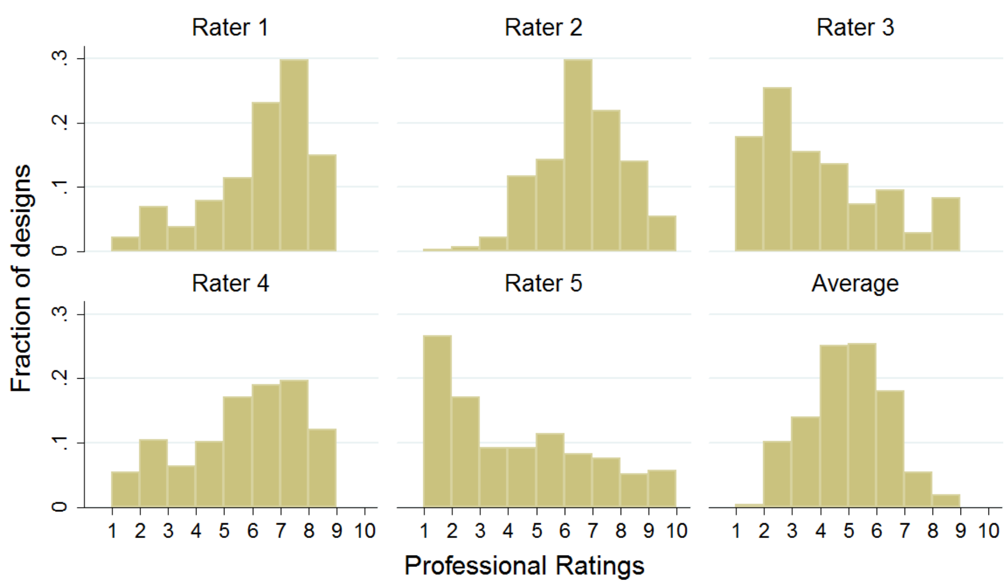
Notes: Table shows the effects of feedback on players' experimentation. Observations are designs. Dependent variable is a continuous measure of the *change* in designs' similarity to the highest-rated preceding entry by the same player, taking values in [-1,1], where a value of 0 indicates that the player's 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 experimenting to copying (and a value of -1, the converse). The mean value of this variable in the sample is -0.01 (s.d. 0.25). Columns (4) to (6) control for time of submission and number of previous designs entered by the player and competitors. Similarity scores in this table are calculated using a difference hash algorithm. Preceding designs/ratings are defined to be those entered/provided at least 60 minutes prior to the given design. *, **, *** represent significance at the 0.1, 0.05, and 0.01 levels, respectively. SEs clustered by player in parentheses.

F Collection of Professional Ratings

The panelists that participated in the ratings exercise were recruited through the author's personal and professional networks and hired at their freelance rates. All have formal training and experience in commercial graphic design, and they represent a diverse swath of the profession: three panelists work at advertising agencies, and two others are employed in-house for a client and primarily as a freelancer (respectively).

Ratings were collected though a web-based application created and managed on Amazon Mechanical Turk. Designs were presented in random order and panelists were limited to 100 ratings per day. With each design, the panelist was provided the project title and client industry (as they appear in the design brief in the source data) and instructed to rate the "quality and appropriateness" of the given logo on a scale of 1 to 10. Panelists were asked to rate each logo "objectively, on its own merits" and not to "rate logos relative to others." Figure F.1 provides the distribution of ratings from each of the five panelists and their average.

Figure F.1: Panelists' ratings on subsample of sponsors' top-rated designs



Notes: Figure shows the distribution of professionals' ratings on all 316 designs in the dataset that received the top rating from contest sponsors. Professional graphic designers were hired at regular rates to participate in this task. Each professional designer provided independent ratings on every design in the sample rated 5 stars by a contest sponsor. Ratings were solicited on a scale of 1-10, in random order, with a limit of 100 ratings per day.

It can be seen in the figure that one panelist ("Rater 5") amassed over a quarter of her ratings at the lower bound, raising questions about the reliability of these assessments: it is unclear what the panelist intended with these ratings, why such a high proportion was given the lowest rating, and whether the panelist would have chosen an even lower rating had the option been available. The panelist's tendency to assign very low ratings became apparent after the first day of her participation, and in light of the anomaly, the decision to omit this panelist's ratings from the analysis was made at that time. The results of the paper are nevertheless robust to including ratings from this panelist that lie above the lower bound.

G Discussion of Social Welfare

Continued experimentation is always in the sponsor's best interest. But the implications for players' welfare and social welfare are ambiguous: the social benefits to innovation can exceed the private benefits, and the social costs will always be greater than the individual designer's cost, due to the negative externalities from competitive business stealing. In this appendix, I elaborate on the welfare implications of prize competition as a mechanism for procuring innovation, focusing on the model of Section 1.

Whether a player's effort is socially optimal depends on the incremental value it generates and the cost of the effort incurred. By this criterion, even tweaks can be desirable, since they come with a new draw of the stochastic component (ε) of the innovation's value. To formalize the argument, let V_{jt} be the value of the most valuable design to-date prior to the t-th design by player j, and let $\nu_{jt} = \ln(\beta_{jt}) + \varepsilon_{jt}$ continue to denote the value of design jt, as in the body of the paper. A new design will only be socially optimal if it is higher-value than V_{jt}, which occurs with probability $Pr(\nu_{jt} > V_{jt})$; otherwise, it will be discarded. Letting Π^S denote social welfare, we can write the expected welfare gains as follows:

$$\begin{aligned} E\left[\Delta\Pi^S\right] &= \underbrace{E\left[\nu_{jt} - V_{jt} \mid \nu_{jt} > V_{jt}\right] \cdot Pr\left(\nu_{jt} > V_{jt}\right)}_{\text{Expected incremental value of an upgrade}} + \underbrace{0 \cdot Pr\left(\nu_{jt} \leq V_{jt}\right)}_{\text{Design discarded}} - \text{cost of effort} \\ &= E\left[\ln(\beta_{jt}) + \varepsilon_{jt} - V_{jt} \mid \varepsilon_{jt} > V_{jt} - \ln(\beta_{jt})\right] \cdot Pr\left(\varepsilon_{jt} > V_{jt} - \ln(\beta_{jt})\right) - \text{cost}\ , \end{aligned}$$

which may be greater than or less than zero. Note that this expression omits the change in each players' expected earnings, which offset each other – the net effect is strictly a function of the beneficiary's gains and the player's private costs. The condition for a socially optimal decision-rule thus reduces to whether the innovation value exceeds the private cost of innovating, be it radical or incremental.

Private choices can deviate from the social optimum under a multitude of circumstances. Because the private benefits are bounded at the dollar prize, whereas the social benefits are unbounded – and potentially quite large, if the fruits of innovation are enjoyed by an entire society – innovation can be inefficiently low unless the prize fully reflects the social value of the innovation. This is more likely to occur in explicit tournaments than in market settings, where the prize is monopoly rents, the size of which are dynamically determined by the level and shape of demand. On the other hand, rent-seeking motives may encourage players to exert effort that increases their expected earnings but yields no net value. A more precise understanding of social welfare would require a specific empirical example or parametrization.

H Summary of Companion Paper: Gross (2014)

In a companion paper, I use a sample of 4,294 logo design contests from the same setting to study the effects of feedback on tournament outcomes. Whereas the present paper studies the effects of feedback on the creative process, in the companion paper I focus on the effects of feedback on the quality of creative outcomes; the distinction is that of the process of innovation versus its result, which in this case is a copyrightable product. Gross (2014) shows that feedback affects design quality via two channels: a selection effect, whereby unsuccessful players are driven to exit and successful players continue, and a direction effect, which guides continuing players towards better designs. In the paper, I use a combination of structural estimation and counterfactual simulations to establish that improvements in quality resulting from feedback accrue entirely as a result of direction rather than selection. These findings imply that successful innovation in this setting requires continuous learning and improvement substantially more than talent or luck.

To highlight some of the basic features and relationships in the tournaments studied here, Section 2 of the present paper reproduces a subset of the results in Gross (2014). In particular, Table 4 uses the companion paper's sample to estimate the relationship between contest characteristics such as the prize, difficulty, and frequency of feedback and key outcomes, and Table 5 uses it to estimate the relationship between a design's rating and its probability of being selected as the winner. I invoke the Gross (2014) sample in these cases because they require a large sample of contests to obtain precise, consistent estimates, and because they are broadly similar across a large set of observable characteristics, as shown below.

The dataset in Gross (2014) consists of nearly all logo design contests with open bidding completed on the platform between July 1, 2010 and June 30, 2012, excluding those with zero prizes, multiple prizes, mid-contest rule changes, or otherwise unusual behavior, and it includes nearly all of the same information as the sample in this paper – except for the designs themselves. Although this sample comes from a slightly earlier time period than the one in the present paper (which was collected in the fall of 2013), both cover periods well after the platform was created and growth had begun to stabilize.

Table H.1 compares characteristics of contests in the two samples. The contests in the Gross (2014) sample period are on average slightly longer, offer larger prizes, and attract a bit more participation relative to the sample of the present paper, but otherwise, the two samples are similar on observables. These differences are mostly due to the presence of a handful of outlying large contests in the Gross (2014) data. Interestingly, although the total number of designs is on average higher in the Gross (2014) sample, the number of designs of each rating is on average the same; the difference in total designs is fully accounted for by an increase in unrated entries. The most notable difference between the two samples is in the fraction of contests with a committed prize (23 percent vs. 56 percent). This discrepancy is explained by the fact that prize commitment only became an option on the platform halfway through the Gross (2014) sample period. Interestingly, the fraction of contests awarded is nevertheless nearly the same in the two samples.

Table H.1: Comparing Samples: Contest characteristics

	Gross (2014)	This paper
Sample size	*4,294*	*122*
Contest length (days)	9.15	8.52
Prize value (US$)	295.22	247.57
No. of players	37.28	33.20
No. of designs	115.52	96.38
5-star designs	3.41	2.59
4-star designs	13.84	12.28
3-star designs	22.16	22.16
2-star designs	16.04	17.61
1-star designs	10.94	12.11
Unrated designs	49.14	29.62
Number rated	66.38	66.75
Fraction rated	0.56	0.64
Prize committed	0.23	0.56
Prize awarded	0.89	0.85

Tables H.2 and H.3 compare the distribution of ratings and batches in the two samples. The tables demonstrate that individual behavior is consistent across samples: sponsors assign each rating, and players enter designs, at roughly the same frequency. The main differences between the two samples are thus isolated to a handful of the overall contest characteristics highlighted in the previous table.

Table H.2: Comparing Samples: Distribution of ratings

	Gross (2014)	This paper
Sample size	*285,052*	*8,144*
1 star (in percent)	16.48	18.15
2 stars	24.16	26.39
3 stars	33.38	33.19
4 stars	20.84	18.39
5 stars	5.13	3.88
	100.00	100.00

Table H.3: Comparing Samples: Design batches by size of batch

	Gross (2014)	This paper
Sample size	*335,016*	*8,072*
1 design (in percent)	72.46	71.84
2 designs	17.04	18.62
3 designs	5.75	5.57
4 designs	2.50	2.19
5+ designs	2.25	1.77
	100.00	100.00