



University of Pennsylvania
ScholarlyCommons

Management Papers

Wharton Faculty Research

2015

The Impact of Visibility in Innovation Tournaments: Evidence From Field Experiments

Joel O. Wooten
University of Pennsylvania

Karl T. Ulrich
University of Pennsylvania

Follow this and additional works at: https://repository.upenn.edu/mgmt_papers

 Part of the [Management Sciences and Quantitative Methods Commons](#)

Recommended Citation

Wooten, J. O., & Ulrich, K. T. (2015). The Impact of Visibility in Innovation Tournaments: Evidence From Field Experiments. <http://dx.doi.org/10.2139/ssrn.2214952>

This paper is posted at ScholarlyCommons. https://repository.upenn.edu/mgmt_papers/231
For more information, please contact repository@pobox.upenn.edu.

The Impact of Visibility in Innovation Tournaments: Evidence From Field Experiments

Abstract

Contests have a long history of driving innovation, and web-based information technology has opened up new possibilities for managing tournaments. One such possibility is the visibility of entries – some web-based platforms now allow participants to observe others’ submissions while the contest is live. Seeing other entries could broaden or limit idea exploration, redirect or anchor searches, or inspire or stifle creativity. Using a unique data set from a series of field experiments, we examine whether entry visibility helps or hurts innovation contest outcomes and (in the process) also address the common problem of how to deal with opt-in participation. Our eight contests resulted in 665 contest entries for which we have 11,380 quality ratings. Based on analysis of this data set and additional observational data, we provide evidence that entry visibility influences the outcome of tournaments via two pathways: (1) changing the likelihood of entry from an agent and (2) shifting the quality characteristics of entries. For the first, we show that entry visibility generates more entries by increasing the number of participants. For the second, we find the effect of entry visibility depends on the setting. Seeing other entries results in more similar submissions early in a contest. For single-entry participants, entry quality “ratchets up” with the best entry previously submitted by other contestants if that entry is visible, while moving in the opposite direction if it’s not. However, for participants who submit more than once, those with better prior submissions improve more when they cannot see the work of others. The variance in quality of entries also increases when entries are not visible, usually a desirable property of tournament submissions.

Keywords

innovation, tournament, contest, open innovation, crowdsourcing, visibility, blind, unblind, idea generation, brainstorming, learning, problem solving, creativity, ideation

Disciplines

Management Sciences and Quantitative Methods

The Impact of Visibility in Innovation Tournaments: Evidence from Field Experiments

Joel O. Wooten and Karl T. Ulrich

Moore School of Business, University of South Carolina, Columbia, SC 29208
The Wharton School, University of Pennsylvania, Philadelphia, PA 19104
joel.wooten@moore.sc.edu, ulrich@wharton.upenn.edu

Version of February 12, 2015

Working paper available from SSRN via <http://ssrn.com/abstract=2214952>

Contests have a long history of driving innovation, and web-based information technology has opened up new possibilities for managing tournaments. One such possibility is the visibility of entries – some web-based platforms now allow participants to observe others’ submissions while the contest is live. Seeing other entries could broaden or limit idea exploration, redirect or anchor searches, or inspire or stifle creativity. Using a unique data set from a series of field experiments, we examine whether entry visibility helps or hurts innovation contest outcomes and (in the process) also address the common problem of how to deal with opt-in participation. Our eight contests resulted in 665 contest entries for which we have 11,380 quality ratings. Based on analysis of this data set and additional observational data, we provide evidence that entry visibility influences the outcome of tournaments via two pathways: (1) changing the likelihood of entry from an agent and (2) shifting the quality characteristics of entries. For the first, we show that entry visibility generates more entries by increasing the number of participants. For the second, we find the effect of entry visibility depends on the setting. Seeing other entries results in more similar submissions early in a contest. For single-entry participants, entry quality “ratchets up” with the best entry previously submitted by other contestants if that entry is visible, while moving in the opposite direction if it’s not. However, for participants who submit more than once, those with better prior submissions improve more when they cannot see the work of others. The variance in quality of entries also increases when entries are not visible, usually a desirable property of tournament submissions.

Keywords: innovation, tournament, contest, open innovation, crowdsourcing, visibility, blind, unblind, idea generation, brainstorming, learning, problem solving, creativity, ideation

Acknowledgments: The Mack Center for Technological Innovation at The Wharton School provided financial support for this research; the Wharton Behavioral Lab provided assistance with data gathering.

1 Introduction

The key to a successful innovation tournament lies in the ability to extract the best *few* opportunities from a process that considers *many* (Terwiesch and Ulrich 2009). In such contests, participation by many agents can reduce individual effort and investment thanks to negative economic incentives (Taylor 1995, Fullerton and McAfee 1999, Che and Gale 2003) but these costs are offset by gains from the parallel search efforts of the increased number of contestants (Terwiesch and Xu 2008, Boudreau et al. 2011). This important characteristic has made tournaments effective processes for generating high quality solutions to innovation challenges (Terwiesch and Xu 2008, Terwiesch and Ulrich 2009). However, when faced with designing such contests, administrators face numerous decisions with respect to how the contest will run – from defining the challenge to soliciting entries to moderating the contest. Knowing that participants adapt to different incentives and information, a key managerial challenge is how a contest administrator can best design and operate a tournament.

In this paper, we examine the effectiveness of two methods of moderating entries to a contest – *blind* and *unblind*. In blind contests, an entry’s visibility is limited to the individual who submitted it and the contest administrator. Without observing the work of others, agents must innovate on their own *from scratch*. In unblind contests, entries are fully visible to other participants; anyone can see the full slate of submissions. The ability to observe directly some positions in the space of possibilities means that agents no longer operate in a vacuum. Seeing other entries could broaden or limit idea exploration, redirect or anchor searches, or inspire or stifle creativity. What effect does entry visibility have on contest performance?

To answer this question, we report on a set of field experiments using web-based platforms for graphic design tournaments. We manipulate contest visibility – either blind or unblind – and use real contests and designers to test how changing the information available in the search process impacts exploration. Specifically, our goal is to test for differences in participant behavior and contest outcomes that stem from the administrator’s decision about entry visibility. As a secondary benefit, we also address the common problem of how to deal with opt-in participation, which often occurs in the context of open innovation and innovation tournaments.

Our experiment is unique in that it is the first to look at differences between innovation tournaments with varying degrees of entry visibility. The eight contests we launched resulted in 665 submissions from 224 agents over the course of a week. A panel of target consumers then rated the quality of each entry, giving us 11,380 distinct entry-ratings. Additionally, students grouped the entries into related clusters in order to quantify the similarity between submissions. These measures – along with the detailed contest administration data – allow us to analyze both participant entry and the characteristics of the work these participants submit.

Our results show that there are, in fact, differences in agent behavior and contest outcomes based on the degree of entry visibility. We find that unblind tournaments generate more entries – not by inducing more entries from existing agents but by increasing the number of agents that participate. We also find that the degree of similarity among submissions increases in early periods, provided that agents can see other entries. For single-entry participants, entry quality “ratchets up” with the best entry submitted by other contestants previously if that entry is visible, while moving in the opposite direction if it’s not. Unblind contests offer an environment in which to learn about the landscape and produce better entries. However, for participants who submit more than once, those with better prior submissions improve more when they cannot see the work of others. The variance in quality of entries also increases when entries are not visible, usually a desirable property of tournament submissions.

2 Visibility in Innovation Tournaments

Innovation tournaments have been shown to be effective processes for generating novel solutions (Terwiesch and Xu 2008, Terwiesch and Ulrich 2009). In fact, they have a long history of driving progress, especially in the fields of engineering and design. Consider the famous Tower Bridge in London, the largest and most sophisticated bascule and suspension bridge ever constructed when it went up. At that time, London’s East End faced massive congestion, and delays for pedestrians and vehicles were routinely several hours. A “Special Bridge or Subway Committee” convened in 1876 and announced a contest to design a new public crossing on the Thames that wouldn’t disrupt commercial river traffic; over 50 designs were submitted for consideration and produced the final design we see today.¹

Such tournaments have typically been organized as blind contests with batched evaluation. That is, designers submit one or more entries – without knowing what other ideas are submitted – and wait for a panel to declare a winner. More recently, developments in information technology in several domains have made submission and evaluation of entries to tournaments much less costly, allowing for sequential in-process evaluation. For instance, the 2006 Netflix Prize sought a new recommendation algorithm for its movie business. By automating the judging, Netflix could provide instantaneous scoring of submissions, allowing the 5,169 teams (who submitted over 44,000 algorithms in total) to learn the quality of their entries throughout the contest and resubmit.² Netflix employed a blind contest with sequential evaluation – entrants still couldn’t see the ideas that were submitted, but were scored in real-time and shown the distribution of results. Sequential scoring effectively changes innovation tournaments from one-shot events to dynamic environments in which individuals can participate and learn iteratively.

Information technology has enabled other modifications to traditional contest features. One such element is the blind constraint. Rather than maintain the precedent of restricting entry visibility, some platforms have pulled back the curtain, allowing entrants to see the work of other contestants. This raises

¹ The Corporation of London and the Tower Bridge Exhibition, www.towerbridge.org.uk (2013)

² Netflix Prize, www.netflixprize.com (2013)

the question of whether seeing other submissions helps or hurts contest outcomes. Anecdotally, the market believes visibility of entries influences outcomes. The web-based contest platform 99Designs, one of the sites we use in our field experiments, advertises that blind contests attract better designers, promote creativity, and result in higher quality entries.³ Contrary to the popular adage echoed by Isaac Newton in a 1676 letter – “If I have seen further it is by standing on the shoulders of giants” – this claim suggests that seeing prior solution attempts does not improve outcomes. However, there has been no prior empirical evidence that has directly explored this impact of entry visibility in innovation tournaments.

3 Literature and Hypothesis Development

Before we develop our hypotheses, we categorize some of the current literature in terms of the type of contest examined (Table 1). This classification is used throughout the rest of the paper and, more generally, as a review of the research that directly deals with innovation contests.

We examine the impact of moderating entry visibility by looking at blind and unblind contests. In this paper, *visibility* specifically refers to the state of transparency surrounding an entry’s full and complete solution. In blind contests, an entry’s visibility is limited to the individual who submitted it and to the contest administrator; other participants may see ancillary information – such as who submitted it or the rating it received – but *not* the innovation itself. This requires agents to innovate on their own. In unblind contests, submitted entries are fully visible to other participants; anyone can see the full slate of submissions.

In what ways might visibility of entries alter tournament outcomes? Once an agent has committed to join a tournament, visibility of other entries should theoretically be beneficial to his or her problem solving efforts. The other entries can be viewed simply as additional information – and from that perspective should not degrade an agent’s performance relative to not having that information. Indeed, an agent could simply ignore the other entries and work from the problem statement with no other information. The agent could then consider the other entries, and decide whether or not to create additional entries based on that newly available information.

While theoretically appealing, this argument may not reflect the realities of human behavior. People are unlikely to actually ignore readily visible entries from rivals, especially as they consider whether or not to join a tournament. Thus, the visibility of the entries of others is likely to influence the outcome of a tournament in at least two basic ways. First, the visibility of the entries of others may influence the likelihood of entry from an agent, altering the number of entrants, their composition, and number of entries each submits. Second, the visibility of the entries of others may influence the way in which a particular agent addresses the challenge, possibly leading to differences in the search process and quality of entries submitted by that agent. We refer to these two pathways of influence as *entry* and

³ 99Designs, www.99designs.com (2013)

Table 1. Tournament Characteristics within the Literature

Paper	Methodology	Visibility	Evaluation	Key Result
Wooten & Ulrich (2014)	Field experiment (99Designs, CrowdSpring)	Unblind	Sequential	Seeing feedback increases entrants and quality; Quality ratchets up with prior best
Bayus (2012)	Observational study (IdeaStorm)	Unblind	Sequential	Ideators can get stuck, generating less diverse ideas after success
Erat & Krishnan (2012)	Analytical	Blind	Batched	Searchers cluster in regions (search breadth increases with entrants sub-linearly)
Poetz & Schrier (2012)	Natural experiment (company)	Blind	Batched	Users in the market can outperform professionals in some ideation settings
Bockstedt et al. (2011)	Observational study (LogoMyWay)	Unblind	Sequential	Winning agents enter earlier, submit over time (but aren't helped by add'l entries)
Boudreau et al. (2011)	Observational study (TopCoder)	Blind	Sequential	Uncertainty amplifies the benefit of parallel search
Kornish & Ulrich (2011)	Lab experiment (university)	Blind	Batched	Incidence of redundancy in parallel search is small (<13% across several domains)
Walter & Back (2011)	Observational study (Atizo)	Unblind	Batched	Increases in rewards and market maturity attract more submissions
Bullinger et al. (2010)	Observational study (university)	Unblind	Sequential	Cooperation amongst agents can be a beneficial search strategy
Jeppesen & Lakhani (2010)	Observational study (InnoCentive)	Blind	Batched	Agents from outside fields have a higher chance of winning
Y. Yang et al. (2010)	Observational study (TaskCN)	Blind	Sequential	Higher award levels, less crowded markets, and easier tasks lead to more entries
Archak & Sundararajan (2009)	Analytical	Blind	Batched	All significant outcomes will be determined by the most efficient contestant types
J. Yang et al. (2008)	Observational study (TaskCN)	Blind	Sequential	Average users do not increase their chance of winning with experience
Terwiesch & Xu (2008)	Analytical	Blind	Batched	More entrants and parallel search offset underinvestment by any one contest agent
Lakhani et al. (2007)	Observational study (InnoCentive)	Blind	Batched	Winning agents work twice as long on their solutions
Soukhoroukova et al. (2007)	Field experiment (company)	Unblind	Sequential	Group decisions in visible idea markets can improve evaluation
Toubia & Flores (2007)	Sim, Field experiment (company)	Blind	Batched	Ideas can be misclassified or misjudged by users
Dahl & Moreau (2002)	Lab experiment (university)	Blind	Batched	Analogical thinking can benefit ideation strategies
Gradstein & Konrad (1999)	Analytical	Blind	Batched	Multi-stage contests differ from one-shot events

characteristics of entries, respectively. We consider each of these in turn, relating the effects to the literature and posing hypotheses for our experiments.

3.1 Entry

The number of entries to a tournament is a function of both the number of entrants to the tournament and the number of entries submitted by each entrant. Here we consider how entry visibility impacts each of these variables.

To begin, we acknowledge the long-established incentive effect in contests, which equates to agents reducing their effort or participation in response to increased numbers of competitors. This relationship has gathered considerable attention in the economics literature (Lazear and Rosen 1981, Casas-Arce and Martínez-Jerez 2009) and innovation contest literature (Taylor 1995, Fullerton and McAfee 1999, Che and Gale 2003). Recently, researchers have emphasized how distinct features in the context of innovation problems can overcome these negative contest incentives (Terwiesch and Xu 2008, Boudreau et al. 2011). Boudreau et al. (2011) demonstrate that increased uncertainty – in terms of who will achieve the best outcome – reduces the incentive to scale back effort. Based on this prior result, a first prediction might assert that increasing entry visibility should reduce uncertainty and therefore reduce participation. However, we will make the opposite claim.

The view that entry visibility will reduce uncertainty rests on the implicit idea that seeing a full solution provides information about its probability of winning. It is important to note that ancillary information – such as who submitted an entry or the rating it received – may be observed independent of the state of entry visibility. A large part (or even all) of the information to be gleaned about the probabilities of winning may reside with such ancillaries instead of the solution specification. More importantly, however, visible entries reveal a different type of information. Seeing the ideas of others can inform participants about the landscape of possible solutions and the administrator’s quality function. Before running with the ‘uncertainty effect’ prediction above, we examine the implications of this more nuanced effect of landscape and learning. The number of contest entrants could increase with entry visibility because of a lower cost of entry, more appealing community experience, and from a superiority bias on the part of entrants, or the number of entrants could decrease as a result of intellectual property concerns.

Individuals might face lower entry costs thanks to having a better map of the solution landscape, lots of seed ideas from which to begin their search, or exemplars that can be changed incrementally with less work than starting from scratch. In searching for solutions, effective strategies can include analogical thinking (Dahl and Moreau 2002), recombination of acquired expertise (Lakhani et al. 2007), and cooperation among agents competing in the same search (Bullinger et al. 2010). This idea is partly formalized as the path of least resistance, an idea within psychology’s *structured imagination* construct where people modify existing solutions when faced with problems requiring creativity (Ward 1994). We

see one derivation of this idea from Wooten and Ulrich (2014), in which knowing where good ideas occur on the landscape – through visible feedback – results in more contest entrants over time.

Increased visibility could promote more appealing social engagement and intrinsically more interesting work. Seeing other entries could also trigger cognitive biases (Alick et al. 1995) and induce greater participation from a better-than-average self-perception. The overconfidence bias is well documented, with several decades of research in a variety of contexts (Kahneman and Tversky 1977), including those seeking to invent new ideas (Åstebro et al. 2007). Each of these suggests that seeing other entries may result in more entrants per contest.

Bockstedt et al. (2011) highlight one disadvantage of entry visibility; namely, the perceived potential for intellectual property loss. If the perceived threat of having an idea “stolen” is high enough, it could be a deterrent to entry. Of course, instead of opting out, agents could decide to devote more effort and stake a claim to the area around an idea, with increased submissions to discourage infringement from competitors, which leads to our second participation variable – entries per entrant.

The number of entries per entrant could increase with entry visibility thanks to lower search costs – in much the same way as the entry decision could be affected. It is possible that a spirit of competition is induced by revealing the work of the participants. In unblind contests, several empirical studies analyze how contest characteristics impact contestant participation (Table 1), including increased entries with market maturity (Walter and Back 2011) and less complex tasks (Yang et al. 2010). However, most of these studies study total contest entries instead of the behavior of contestants within a contest. Bockstedt et al. (2011) empirically demonstrate that winning agents on LogoMyWay.com are more likely to enter earlier and submit entries over a wider range of time, but aren’t helped by simply entering more ideas.

On balance, we expect that greater entry visibility in innovation tournaments will result in increased participation. All but one of the hypothesized effects suggest that contest entries will be greater in contests with entry visibility. Unblind contests make an agent’s key decisions easier. The choices around whether to enter and the amount of effort to invest both derive benefits from entry visibility. By seeing other entries in the landscape of possibilities, the barriers to entry are lower for any given agent and more information on the administrator’s quality function is available. Easier search should result in more entry.

Hypothesis 1: Increasing entry visibility in an innovation tournament (by moving from a blind to an unblind contest) will increase the number of entries submitted.

3.2 Characteristics of Entries

Given that we expect the number of entries to change, do the characteristics of those entries also change? Entry visibility may influence the way in which a particular agent addresses the challenge, possibly leading to differences in the search process and quality of entries submitted by that agent. Two relevant metrics of the characteristics of entries are *similarity* and *quality*, including both the mean and distribution.

Similarity. Independent of idea quality, seekers usually benefit from knowing the landscape – observing diverse ideas gives a more complete picture of the solution possibilities. The incidence of redundancy in parallel search has been shown to be quite small in blind contests (Kornish and Ulrich 2011). In unblind contests, entry visibility could mean even less redundancy in effort, with agents using the knowledge of competitors’ submissions to reduce repetition. Or such visibility could inhibit parallel search, with entrants clustering their submissions around existing proven entries (Erat and Krishnan 2012). Either way, if a participant searches differently in response to seeing other entries, then the resulting similarity among entries should change.

In a set of graphic design prototyping experiments around online ads, participants who saw multiple shared designs borrowed significantly more features to incorporate in their own ads (Dow et al. 2012). In creativity tasks, Marsh et al. (1996) found that individuals who saw many examples tended to incorporate critical elements in their own designs (although without inhibiting creativity), and Smith et al. (1993) found conformity in every group that saw examples, across a range of conditions and instructions. In unblind contests, more designs will be visible to agents, and we expect the prior conformity results to play out in innovation contests.

Hypothesis 2: Increasing the visibility of entries in an innovation tournament will result in submissions that are more similar.

Quality. At the level of the contest, the population of entries yields a distribution of quality, reflecting the overall performance of the tournament. This idea arises from the statistical view of innovation processes (March 1991; Dahan and Mendelson 2001; Terwiesch and Ulrich 2009). One way to describe the quality distribution is with mean and variance, and increases in each of these variables improve the overall performance of tournaments (Girotra et al. 2010).

The mean quality of entries is driven by both the quality of entrants and the quality of the work those entrants do. If a tournament attracts better entrants or better submissions from its existing entrants, overall contest performance improves. However, in many settings, it’s not possible to truly disentangle the intrinsic quality of entrants from the work they do. Here, we rely on entry quality as the aggregate measure of these two drivers and explore how that quality might be influenced by entry visibility.

Exposure to additional information in unblind contests likely impacts the learning environment. Openness and information sharing has long been identified as important to scientific progress (Merton 1942, Mulkey 1975), with examples such as open source software development at the recent forefront (von Hippel 2005). In evolutionary economics, the role of search has been highlighted as a mechanism for discovering variety and allowing organizations to develop new technologies (Nelson and Winter 1982). Metcalfe (1994) suggests that exploring such variety allows firms to innovate more successfully by seeing a range of potential options or paths to explore.

We would expect participants to learn the most and have the best understanding of the search landscape when full information from all the parallel searches is visible. In the design world, having

examples readily available has been shown to improve the appeal of designs, although these benefits appear to accrue to novice designers more than to experts (Lee et al. 2010). With visible entries, the learning effect could also help with self-filtering. If contestants have some ability to assess their own work, they may balk at submitting lower-quality designs after seeing those already submitted, which would increase mean quality.

Unlike the above arguments, some operators of web-based platforms for innovation contests assert that blind contests result in better entries, with the rationale that blind contests attract higher quality talent. If better designers don't benefit from the presence of examples as Lee et al. find, then other benefits of the blinded format (such as intellectual property protection) could be attractive. On balance, however, we believe that there is more evidence on the side of increased information and learning, as mechanisms for increasing the average quality of entries.

Hypothesis 3: Increasing the visibility of entries in an innovation tournament will increase the average quality of entries submitted.

Finally, variance in the quality of submissions, for a given mean, improves tournament outcomes, as flatter distributions result in more ideas in the upper tail of the distribution (Girotra et al. 2010). Such benefits could be driven by both variance in the quality of entrants and by variance in the quality of the work they do. Given the uncertainty in the task and conditional on a given set of entrants, variance in approach is expected to be one of the key drivers of variance in quality. The way in which an agent searches the landscape likely impacts variance in the quality distribution. Thus, it follows from our similarity hypothesis (H2) that we expect less variance in approach in visible tournaments, and by implication less variation in quality. Wooten and Ulrich (2014) similarly found that more information about the administrator's quality function results in a convergence of approaches and decreased variance in the quality of contest submissions.

Hypothesis 4: Increasing the visibility of entries in an innovation tournament will decrease the variance in quality of entries.

4 Experimental Design

We conducted a set of field experiments in which we explicitly control the environment and compare the performance of innovation contests with varying levels of visibility. We've used these platforms for experiments before; however, here we use a completely new set of experiments designed specifically to address the issue of visibility in contests. We follow similar conventions as those used by Wooten and Ulrich (2014) for the setup, delivery of feedback, and measure of entry quality in an online graphic design field experiment. Four pairs of logo design competitions were posted on two online design contest marketplaces, 99Designs and CrowdSpring. The competitions differed in terms of the amount of information visible to entrants – in the unblind treatment, agents could see all entries and feedback while

in the blind treatments, the entries of others were not visible. At the conclusion of the contests, a consumer panel rated the quality of each entry and a pool of university students rated their similarity.

4.1 Contest Platforms

Our experiments were hosted by two online companies, 99Designs and CrowdSpring, that have emerged as leaders in the crowd-sourced design market. They allow buyers to solicit projects – such as logo creation – from a community of graphic designers. While buyers are mostly small businesses and entrepreneurs, established companies such as Amazon, Starbucks, Microsoft, Philips, Barilla, and TiVo have also run contests. Contest winners are awarded predetermined cash prizes – normally between \$150 and \$1,500 per contest. The sites support robust marketplaces. As an example, 99Designs has awarded over \$79 million worth of contest prizes in more than 317,000 contests since its founding in 2007.

The two platforms are very similar, with nearly identical interfaces and business implementations. Each website counts over 125,000 designers as members and targets an array of design projects (such as logos, packaging, book covers, and website design). Clients create a contest by posting project specifications and a prize amount. Over a project’s duration (generally one week), online submissions are submitted by interested designers and feedback can be given by the client.

4.2 Contests

Four pairs of contests were launched as follows.

A: Burning Barn BBQ Sauce Smoking Silo Salsa	B: Wave Monkey Headphones Sound Chimp Speakers
C: Power Perk Coffee Bold Brew Tea	D: Jailbird Dog Gear Rat Pack Cat Company

All eight contests had similar details, and within each pair, projects had nearly identical details, including company type, name, design specifications, deliverables, target markets, and specifics of the design brief. Each logo was for a new consumer product brand whose target audience was specified to be college-educated U.S. consumers 18-35 years old. Designers were told that a panel of consumers from this market would be the ultimate judges of entry quality. The contests in each pair shared the type of product (condiments, audio electronics, beverages, and pet accessories), were constructed with similar name characteristics and motifs, and were randomly assigned to one of the two websites. An example of the submitted design briefs is Appendix A.

Designers count on feedback over the course of contests to determine performance of any particular entry. The established feedback mechanism on both 99Designs and CrowdSpring is a one-to-five star rating, which indicates how much the administrator likes an entry. We provided new entries with feedback every morning using this scale; a three-person panel of independent judges scored each design and their average determined the rating, expressed as a number of stars. The raters fit the target market demographic (consistent with our design brief), and we used two such panels to manage the volume from four concurrent contests. This feedback was intended to be highly correlated with the final ratings which

would eventually be produced by an evaluation by a larger panel of consumers at the conclusion of the contest.

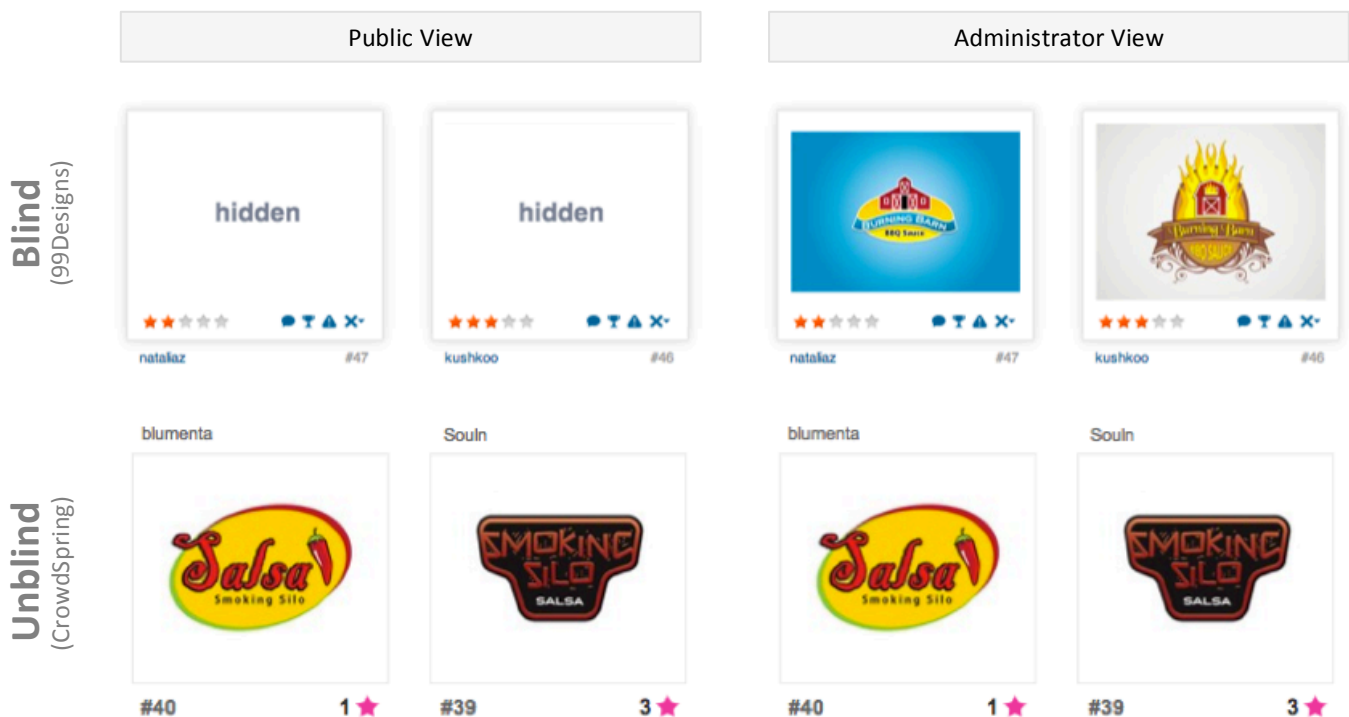
4.3 Treatments

The independent variable tested was entry visibility in each contest. CrowdSpring and 99Designs permit both blind and unblind contests, which allows the administrator to choose at the outset who can see a designer's submissions. In unblind projects, anyone who views the contest can see the full slate of designs that have been entered as well as any scored feedback given (in the form of star ratings). Thus, the general public has full information about submissions and their in-process ratings. In blind projects, an entry's visibility is limited to the designer who submitted it and the contest administrator. Other designers know how many designs have been entered – and by whom – but are restricted from viewing the actual submission. Figure 1 gives examples from set A in our experiments, showing the blind contest views from 99Designs and the unblind views from CrowdSpring as they appeared on the sites.

4.4 Experiment

We denote the four pairs corresponding to the four product types as A, B, C, and D. One of each pair ran on 99Designs and its nearly identical corollary ran on CrowdSpring, allowing for each visibility treatment to be tested twice on each site in a balanced design. Designers closely monitor the contests on these websites and frequently report copyright violations and other such concerns. To deal with such savvy agents and avoid undermining the outcomes, we constructed the experiment design to utilize two different website platforms, slightly staggered start dates, and small differences in the award levels. The

Figure 1. Example of Treatment Information (set A)



contest pairs ran over the course of two weeks. Sets A and B ran during the first week, and sets C and D ran during the second. CrowdSpring and 99Designs display the award amount in slightly different ways, but sets A and C carried award levels of \$250 for the winner, and awards for sets B and D were \$237. These slight differences were built into the contest setup to make the contests nearly identical, without tipping the designers off that the products weren't real. The visibility treatment can be denoted by subscripts (*B* for blind and *U* for unblind) resulting in the following contest layout:

99Designs:	A_B	B_U	C_U	D_B
CrowdSpring:	\bar{A}_U	\bar{B}_B	\bar{C}_B	\bar{D}_U

The eight contests relied on the standard mechanisms of the websites to entice designers to participate; we address the challenges of opt-in participation in section 5. Each contest was open to anyone on the respective website, ran for seven days, received daily feedback, and resulted in an award to the winning designer. All experiments were conducted after obtaining approval from the human subjects committee at the university.

4.5 Evaluation

A total of 665 entries were generated by 224 designers over the course of the eight tournaments. Two panels of 20 judges independently and anonymously evaluated the logos from the perspective of potential consumers. The judges were representative of the target market outlined in the contest briefs – college-educated individuals between the ages of 18 and 35. These judges were similar in profile but distinct from the feedback panelists, who provided the daily star ratings.

Ratings were collected using web-based surveys. One panel of judges rated logos in sets A and D; the other rated logos in sets B and C. Following the design of Wooten and Ulrich (2014), entries from the eight contests were administered in separate surveys and were completed as paired sets. To mitigate order effects, surveys were administered as a balanced, repeated Latin square design; each set order (AD, DA, BC, CB) appeared the same number of times. Within each set, half the judges were given the contest from 99Designs first followed by the one from CrowdSpring; the other half saw them in the opposite order. Within each individual survey, the logos were presented to each judge in a randomized order. The question and response choices (on a 1-5 rating scale) were the same for the judges as for the in-contest feedback panel.

One contest (Wave Monkey Headphones, with 192 entries) exceeded the survey length threshold established in similar settings (Girotra et al. 2010, Wooten and Ulrich 2014). As a result, each judge rated half of the designs in that contest; for those 20 judges, the assignment of particular logos followed a balanced, repeated Latin square design in which each rater saw 96 logos and each logo received 10 raters.

The judges' responses to the eight surveys provide the measure of entry quality for our analysis. We find that the reliability of judges is high (Table 2). We check this using a Krippendorff alpha test on our

Table 2. Inter-rater Reliability of Judges (Krippendorff Alpha)

Contest	<i>Interval Alpha</i> Judges' Ratings	<i>Interval Alpha</i> Bootstrap Averages
A: Burning Barn BBQ Sauce	0.35	0.84
\bar{A} : Smoking Silo Salsa	0.21	0.80
\bar{B} : Sound Chimp Speakers	0.04	0.72
C: Power Perk Coffee	0.12	0.72
\bar{C} : Bold Brew Tea	0.11	0.65
D: Jailbird Dog Gear	0.23	0.74
\bar{D} : Rat Pack Cat Company	0.08	0.57
Average	0.16	0.72

Note: Judges' Ratings analyzes degree of agreement among the 20 judges on every logo's rating; Bootstrap Averages measures the agreement between a 10-judge random sample of the 20-judge panel and the remaining 10 judges on the average logo rating; Wave Monkey Headphones omitted because of Latin square missing values.

population of raters. Given the artistic nature of our contests, we expect high variation in the scores because of personal preferences. This is corroborated with a relatively low degree of agreement between any two judges. However, if populations have stable preferences, then a high degree of agreement should be seen in the average scores of entries across populations. We test this with a bootstrap approach, splitting our judges into two randomized groups and comparing the average scores for each logo between groups. With this population-level approach, we obtain an agreement alpha of 0.72, above accepted thresholds. A sample of the scored logos is provided in Figure 2.

5 The Opt-in Problem

The gold standard in experiment design is randomization, in which subjects are randomly allocated among treatment groups. In our field experiments, subject randomization isn't possible – platform participants freely choose which contests to join. By relying on this platform mechanism, we increase the ecological/external validity of our study but also raise several questions by having subjects opt in. The growing number of online platforms increases the frequency with which these issues are likely to be encountered in empirical studies. We use this section to discuss the opt-in problem in an abstract way and get some purchase on our particular selection issues.

The most basic question when dealing with opting-in (self-selection) in any context is whether the samples mirror the population of interest. In addition, our setup also highlights an increasingly common issue – accounting for participation in a dynamic system. By this, we mean situations in which agents may choose to participate at any point in a defined time window and throughout which characteristics of the opted-into event or system may change. Our contests fit this description. Potential entrants can observe several dynamic variables (including number of entries, specific entrants, and administrator

scoring in both treatments) over the course of each contest. To our knowledge, no one has addressed how to account for opt-in behavior in this setting.

Figure 2. Examples of Logos Generated (sets A and B)

	Highest Rated	Median	Lowest Rated
Burning Barn BBQ Sauce ^{1,B}			
Smoking Silo Salsa ^{2,U}			
Wave Monkey Headphones ^{1,U}			
Sound Chimp Speakers ^{2,B}			

1: 99Designs, 2: CrowdSpring; B: Blind (no visibility), U: Unblind (full visibility)

We examine six scenarios (shown in Table 3) as a framework for addressing opt-in behavior, with the visibility of the treatment and the presence of dynamic information as the conditions of interest. *Treatment visibility* is simply whether the experiment treatment is observable to an agent when they decide to opt in. For example, in our experiments, the ability to see other entries (our treatment) was obvious to potential entrants, so treatment visibility would be coded as ‘yes.’ *Dynamic information* describes whether the information that an agent evaluates in making their opt-in decision can change over time. In our case, certain participation data was publicly known and changed over the week that the contests ran, so dynamic info would also be coded as ‘yes.’ Dynamic systems can be observed in many participation-based markets (i.e., eBay auctions, open innovation contests, crowdfunding projects), making this a growing area of interest.

To ground our discussion, we begin with two situations where participants do *not* opt in. (We explicitly define *opting-in* as a choice of deliberate participation on the part of an agent.) Randomized treatment assignment accounts for both observed and unobserved participant characteristics; this is the best-case scenario, which is often possible in the lab environment. Boudreau et al. (2011) provides a nice example of this case where the composition of randomized individuals in a contest influences the degree of participation. In other situations, a biased or truncated population results in an imperfect sample. Heckman (1979) originally attacked this problem of nonrandomly selected samples via two-step estimation. Correcting endogeneity through econometric techniques has gained considerable attention in the last few decades. Guo and Fraser (2009) offer a compelling overview of sample selection and treatment effect models, dating back to the original econometric framework introduced by Heckman.

Table 3. Methods to Account for Participant Self-Selection

Participant Opt-in?	Treatment Visible?	Dynamic Info?	Correction	Rationale
no (random)	either	either	None	Randomization takes care of both observed and unobserved traits. (Worth checking sample sizes.)
no (biased)	no	no	2-step estimator (Heckman correction)	Adds expl. variable to account for background traits in truncated/biased populations.
yes	no	no	2-step estimator; Effects model	Corrects for sample selection or treatment effects with two steps; Includes fixed/random effects for observable traits to account for samples.
yes	yes	no	None*	Outcomes are of paramount interest. Mechanism interesting but not required. (Could be disentangled via lab study.)
yes	no	yes	} Proposed opt-in characterization hierarchy	Presence of dynamic info permits some visibility into opt-in behavior. Goal is to control dynamic pieces and simulate one of the above scenarios.
yes	yes	yes		

* Depends on type of opt-in; No correction works in certain settings. Observable characteristics could help illuminate participant choices in others. Two-step estimator approach possible.

While labor economics begat many of the now-standard techniques, other fields have begun to adapt their methods and develop richer discussions around endogeneity. As an example, Hamilton and Nickerson (2003) highlight one such exploration in the field of strategic management. We lay this out, not to offer a robust review of the literature, but to signal the opportunity that exists. Here, we offer some cursory thoughts, but a more comprehensive methods paper could attack several of the issues we raise – especially with respect to dynamic information.

In the case of no treatment visibility and no dynamic information (row 3), a 2-step estimator or an effects model could help address the sample opt-ins. When the decision to opt in is influenced by the treatment being observable (row 4), one could argue that no controls or allowances are needed. This could be the case if two treatments draw from the same population and have static information. With no other differences, the treatment becomes part of the opt-in decision. Using no correction focuses on the ultimate outcome – namely, what effect does choosing a particular treatment have? We conjecture that a two-step estimator approach could achieve a workable method (determining the various effects from the participation decision, the treatment effect in the system, and the unobservable traits), but disentangling the unobservables and participation decision is not trivial. A first step might be a switching model for treatment effects, acknowledging that the open choice to opt-in to many contests is not a binary choice. However, using no correction also provides a defensible alternative.

The final two cases (rows 5 and 6) also have dynamic information to contend with. By this, we mean situations in which agents choose their level of involvement based on system observables that change over time. Because the prior cases have no way to deal with dynamic systems, we propose to first address whether the participant populations are similar, then control for the dynamic information influencing opt-ins, and finally use one of the previous solutions (row 3 or 4).

For assessing the populations and dynamic information, some characterization of the population is often needed. Our hierarchy for this characterization of opt-ins consists of three tiers. Tier one has objective measures of a dimension similar to the dependent variable. In our case, we are interested in idea quality, so designer skill, talent, or effort could be the tier one measures. Tier two has measures that are indirectly related to the dependent variable. In our case, experience or education achievement would qualify. Tier three includes other observables that may or may not be directly related with the dependent variable but may contain some information about participants. In our case, designer location or language ability fit those criteria. These tiers are ordered, with tier one likely being more instructive than tier two and so on. In the following sections, we use these tiers to characterize the platform populations and control for the dynamic information in the contests.

5.1 Comparing Platform Populations

In this section, we assess whether the populations on our two platforms are similar. This allows us to compare the contests with more confidence. In some settings, scraping the profile information for every

agent on the platform would accomplish this. In our setting, there are two issues with that approach. First, not all designers are represented by viewable profiles. For example, on CrowdSpring, only 38% of the more than 158,000 designers have an active profile in 2014. Deactivated accounts and site restrictions likely account for most of this drop. (When we add deactivated accounts back in, the average designer reputation score on CrowdSpring falls from 71.5 to 58.9, suggesting that less successful designers leave over time as one would expect.) This finding warrants checking any platform profiling for such omissions. Second, the active population – not the total population – is the group of interest. At the time of the contest, which designers were active on the site and actually had a chance to opt-in to our contest? That more granular population (while sometimes difficult to obtain) makes for a more appropriate comparison. These issues are compounded if the data collection does not happen in parallel with the experiment.

Here, we address both of those concerns by randomly choosing 10 contests on each site (median prize: 99Designs = \$295, CrowdSpring = \$304) from the 3 months immediately after our experiment and using the designers who participated in those 10 contests as a proxy for each platform’s population at the time. We record profile information (from the contest page instead of the possibly-deactivated profile pages) for 296 of 330 unique designers on 99Designs and all 335 unique designers on CrowdSpring. This gives a snapshot of the agents who were active on the sites at the time of our experiment. A summary of their experience and performance is shown in Table 4.

Skill, experience, and demographics are three categories of information that could help assess our two populations. These are ordered, with skill likely being more instructive than experience, which is likely more instructive than demographic info. The top three rows in Table 4 summarize the agents’ experience; the bottom three rows address their skill. We omit demographics since we have data from the more instructive tiers.

Days of platform experience (2011) gives the number of days between a designer registering an account and our capture period concluding (in July 2011). The alternate *all* days measure shows the number of days from registration until the most recent login (adjusted partial results shown for CrowdSpring). *Number of contests entered* reveals the number of contests entered by designers on CrowdSpring; it excludes those profiles (~16%) that withdrew from all contests they didn’t win in an attempt to manipulate their profile stats. In terms of designer experience, we see similar breadth of experience at the time of our experiment (mean of 321 vs. 307 days), with some high-volume designers on the top end of the distribution, as expected.

Talent is harder to measure. We report *number of contests won* as a retrospective look at skill (the figures are from 2014, which is better than observing the same figure in 2011 and penalizing new designers). An issue with *contests won* is that it depends on the number of contests available on the site. We observe a lower mean number of wins on CrowdSpring; however, the ratio of contests to designers is

significantly lower (0.27 vs. 1.01) as well, so the average CrowdSpring designer will collect fewer wins by default.

Another measure of talent is available – each site algorithmically assigns a skill score to each participant, which we denote as *Reputation* (99Designs refers to it as Level). As seen in Table 4, these have different scales. 99Designs’ scale starts from 0 and is additive, with better designers moving up the ladder. A double-digit score is reasonably good in their system and usually corresponds to dozens of wins. CrowdSpring’s scale ranges 0-100 but starts from 70, from which new designers move up or down based on their performance. The skill ratings for 99Designs are right-skewed (with most scores clumped at the bottom) and those for CrowdSpring are left-skewed (with most scores clumped at the top). To permit comparisons, we translated each of these measures of talent into an adjusted z-score, using log and flipped log transformations suggested by Tabachnick and Howell (2007) for the two different skewed scales. Perhaps the most useful measure would be win rate (number of wins/number of contests entered), which is only available for CrowdSpring. However, there is anecdotal data that suggests our metrics are analogous across the two sites. Comparing users who participate on both platforms (4 out of 665 designers in the sample), the site metrics for those four show remarkable consistency (Table 5). While only a small sample, the reputation and win data are well aligned and offer support for using them to represent designer talent, the trait in which we are interested.

In summary, our 20 randomly selected contests produced similar populations, with nearly identical numbers of unique designers (330 vs. 335), similar days of experience (321 vs. 307), and a seemingly similar spread of talent. We offer these comparisons as further evidence (along with section 4) of similar populations on each of our platforms at the time of our experiment. Subsequent analysis uses each of the above measures and explores the samples in a more detailed analysis.

Table 4. Ex Post Platform Comparison of Experience and Talent – 99Designs vs. CrowdSpring

	99Designs				CrowdSpring			
	Min	Median	Mean	Max	Min	Median	Mean	Max
Days of platform experience (2011)	5	178	321	1,659	28	214	307	1,192
Days of platform experience (all)	7	1,131	1,088	2,688	--	--	1,033	2,202
No. contests entered	--	--	--	--	1	70	222	2,811
No. contests won	0.0	5.0	18.2	152.0	0.0	1.0	9.2	232.0
Reputation (raw platform metric)	1.0	6.0	8.7	48.0	1.0	72.0	58.9	100.0
Reputation (adjusted z-score)	-1.4	0.2	0.0	2.0	-1.2	-0.1	0.0	2.9

Note: Represents 296 unique 99Designs designers and 335 unique CrowdSpring designers (May-Jul 2011)

Table 5. Platform Comparison using Identical Designers – 99Designs vs. CrowdSpring

	BrandingDesigner		FishDzn		Alaguraj		Apanasara	
	99D	CS	99D	CS	99D	CS	99D	CS
Reputation (raw platform metric)	40	90	7	81	1	28	1	22
Reputation (adjusted z-score)	1.83	0.77	0.30	0.25	-1.40	-0.88	-1.40	-0.95
No. contests won	64	38	5	4	0	0	0	0
Date of site registration	3/22/11	9/22/09	4/29/10	5/20/10	5/8/11	5/8/11	5/26/11	5/17/11

5.2 Assessing Opt-in Samples

Given similar populations, our next issue addresses whether the opt-in samples in our experiment mirror those populations. Table 6 shows our skill metrics (*# Wins* and *Reputation*) for each of the contests. These match up well with our population data, with mean wins (20.7 for 99Designs and 11.1 for CrowdSpring) not differing significantly from the population set means (18.2 and 9.2). The reputation metrics are also aligned (0.0 and 0.1 sample vs. 0.0 and 0.0 population) for the two sites. Additionally, the highlighted rows in Table 6 give a glimpse into the role that dynamic information plays in the opt-in story. The two contests with the best entrants (more wins and better reputation) are also the ones with the fewest participants. This suggests that entry by successful designers may preclude other designers from opting-in. This idea echoes the findings of Boudreau et al. (2011), where the composition of individuals in programming contests influenced the degree of participation. We examine this impact of dynamic information more fully in section 7.

Table 6. Sample Comparison in Experiment

Contest	Blind?	Site	# Entrants	# Entries	# Wins		Reputation	
					Mean	Median	Mean	Median
BurningBarn	Y	99D	36	82	17	9	-0.15	-0.19
WaveMonkey	N	99D	48	192	17	6	-0.05	0.01
PowerPerk	N	99D	30	91	24	6	0.08	0.01
JailbirdDog	Y	99D	8	40	25	18	0.20	0.36
SmokingSilo	N	CS	24	57	6	3	-0.14	-0.18
SoundChimp	Y	CS	20	53	21	7	0.36	0.15
BoldBrew	Y	CS	29	69	10	3	0.01	-0.26
RatPack	N	CS	29	81	8	4	0.21	0.02

Note: Lowest participation (highlighted rows) in contests with best entrants (highest wins and reputation).

6 Data

Table 7a is a summary of each contest's outcome. Given our experimental design, we also have a great deal of entry-level data, which we analyze to test our four hypotheses.

6.1 Measuring Entry

To measure entry behavior we capture *entries*, *entrants*, and *entries per entrant*.

An *entry* is defined as an idea submission to a particular contest and captures the aggregate level of participation in a contest. The more participation a contest elicits, the more entries there are, resulting in more potential solutions for the contest administrator.

An *entrant* is a distinct contest participant, someone who submits at least one entry. The more attractive the contest, the more entrants it attracts, which increases the number of parallel searches that occur.

Entries per entrant is defined as the number of submissions by a contest participant. We use it to estimate the effort invested by an entrant with the idea that submitting more entries requires additional effort.

Table 7a. Contest Summary

	Condiments		Audio Electronics		Beverages		Pet Accessories	
	Burning Barn BBQ Sauce	Smoking Silo Salsa	Wave Monkey Headphones	Sound Chimp Speakers	Power Perk Coffee	Bold Brew Tea	Jailbird Dog Gear	Rat Pack Cat Company
Visibility	Blind	Unblind	Unblind	Blind	Unblind	Blind	Blind	Unblind
Website	99D	CS	99D	CS	99D	CS	99D	CS
N Entries	82	57	192	53	91	69	40	81
N Entrants	36	24	48	20	30	29	8	29
Best Logo	3.75	3.65	3.70	3.26	3.15	3.20	3.35	3.40
Mean Logo	2.54	2.44	2.40	2.32	2.21	2.51	2.36	2.45
S.D.	0.77	0.58	0.50	0.35	0.48	0.45	0.63	0.39
Average Similarity	0.186	0.196	0.217	0.161	0.186	0.164	0.170	0.210

Note: Values listed for logos are averages of judges' ratings on 1-5 scale; Wave Monkey scores adjusted for judge.

6.2 Measuring Characteristics of Entries – Similarity

The second question we ask concerns the search process: How does the ability to see other entries change the way in which agents address the challenge?

To assess whether agents incorporate elements from previously submitted entries, we need a quantitative measure of logo similarity. Kornish and Ulrich (2011) tackle a similar problem in rating sets of innovation opportunities. We adopt a similar methodology in order to obtain a similarity score for every pair of entries in a contest. We had student subjects in the university behavioral laboratory form groups of similar entries from a packet of logo submissions. A packet contained a subset of logos from a

single contest; each logo was printed on a square of cardstock. This allowed students to visually sort and re-sort the logos into piles quickly. Entries could be categorized into more than one cluster.

We created 45 such packets, with overlapping subsets of entries such that most entry pairs appeared multiple times. When multiple logos varied only by color, we only included only one version of a logo. In total, we ran 89 students through our protocol. They were paid \$5 for participating. Sessions were not timed and most students finished the grouping task in 10-20 minutes. The grouping task resulted in a list of idea clusters that we could turn into a measure of pairwise similarity. The average cluster contained 5.5 logos per group. We coded each of the entries grouped together as *similar* and calculated an overall score between every possible pair based on the percentage of times those two entries were placed in the same cluster. This measure is the number of times two logos were grouped together over the number of times such a pairing was possible. The final score is modified to account for our packet structure and the subsets included. The similarity score between any two entries i and j is represented in the matrix A_{ijk} , where k represents the contest.

To measure how changes in entry visibility affect the similarity of submissions, our similarity metric takes two forms. *Average contest similarity* is the mean of all possible pairwise similarities within a contest, \bar{A}_k . *Logo-level similarity* is the similarity score of a particular logo based on all the logos submitted before it. In other words, how similar a particular logo j is to prior logos i , \bar{A}_{ijk} where $i < j$, for each k .

6.3 Measuring Characteristics of Entries – Quality

To measure contest quality, we operationalize the quality distribution through two parameters – mean and variance. One benefit of this approach is that it breaks the measure of quality into underlying variables and helps mitigate the problem of sampling only winning ideas, which can be noisy in a small sample of contests.

For the *quality* models, the quality measure comes from the judges' scoring of contest entries and the unit of analysis is the rating of a particular judge of a particular logo. Individual logo rating is thus the dependent variable with the following independent variables: treatment (Blind), cumulative number of entries (Entries), the best prior submission by others (Max), and the agent's best prior entry (Personal best).

For variance in the quality of ideas, we construct a measure of *variance* for the dependent variable that takes out the linear quality improvement trend over the course of the contests, identical to Wooten and Ulrich (2014) and similar to Girotra et al. (2010). Table 7b provides descriptive statistics and correlations for the variables used in our analysis.

Table 7b. Descriptive Statistics and Correlations

	Variable	Definition
(1)	Rating	Numerical score of quality for an idea from judge
(2)	Visibility (blind)	Contest visibility treatment – 0: Unblind, 1: Blind
(3)	Domain	Control for product area – 0: Condiments, 1: Audio elec., 2: Beverages, 3: Pet prod.
(4)	Site	Control for platform – 0: 99Designs, 1: CrowdSpring
(5)	Day	Control for day of contest – 1: Mon, 2: Tue, ... 7: Sun
(6)	Judge	Control for individual providing the rating
(7)	# Prior entries	Number of entries submitted to a contest at a logo's time of entry
(8)	Max	Highest score produced by others in contest thus far
(9)	Personal best	Highest score produced by a given entry's agent in contest thus far
(10)	# Entries	Number of submissions in a contest (or over a specified time)
(11)	# Entrants	Number of unique participants who submit at least one entry
(12)	# Entries/entrant	Number of entries submitted by each entrant in a contest
(13)	Avg. contest similarity	Mean of all possible pairwise similarity ratings from survey panel in a contest

Logo-level correlations (11,380 observations):

Variable	Mean	St. dev.	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(1) Rating	2.41	1.18								
(2) Visibility (blind)	0.37	0.48	0.03							
(3) Domain	1.40	1.01	-0.03	-0.15						
(4) Site	0.39	0.49	0.02	0.07	0.18					
(5) Day	5.09	1.89	-0.03	0.03	0.00	-0.00				
(6) Judge	22.68	11.34	-0.03	-0.06	0.05	-0.12	0.05			
(7) # Prior entries	52.77	44.66	-0.02	-0.31	-0.1	-0.3	0.58	0.23		
(8) Max	3.04	0.96	0.03	0.06	-0.11	0.05	0.42	-0.18	0.4	
(9) Personal best	1.75	1.32	0.07	-0.07	0.08	-0.08	0.11	0.01	0.12	-0.09

Table 7b (continued)

Contest-level correlations (8 observations):

Variable	Mean	St. dev.	(1)	(2)	(3)	(4)	(10)	(11)	(12)
(1) Rating	2.41	0.04							
(2) Visibility (blind)	0.50	0.19	0.26						
(3) Domain	1.50	0.42	-0.30	0.00					
(4) Site	0.50	0.19	0.22	0.00	0.00				
(10) # Entries	83.13	16.67	0.06	-0.50	-0.18	-0.41			
(11) # Entrants	28.13	4.16	0.31	-0.42	-0.41	-0.24	0.87		
(12) # Entries/entrant	3.06	0.34	-0.33	0.01	0.52	-0.56	0.19	-0.28	
(13) Avg. contest similarity	0.19	0.01	0.16	-0.83	-0.10	-0.18	0.70	0.65	0.07

7 Analysis and Results

7.1 Entry

Our main variable of interest is *entry visibility*, denoted in our experiments as either blind (low visibility) or unblind (high visibility). To understand how differences in entry visibility affect agent behavior, we estimate variations of the model:

$$Y_i = \alpha + \beta(\text{Entry Visibility})_i + \delta_i + \varepsilon_i .$$

The dependent variable Y varies over the contests i and takes on one of the outcome variables discussed above (entries, entrants, entries per entrant). Since these measures are all counts, our model assumes a negative binomial distribution⁴, which adds an over-dispersion parameter and is generally more conservative than estimates with a Poisson count model (Hilbe 2011). To control for differences across contests, which could influence our behavior measures if not accounted for, we include several fixed effect controls (δ_i) for the *domain* and *site*. Table 7a provides variable details and gives descriptive statistics and correlations for the variables used in our analysis.

Table 8 shows the results of our negative binomial regression analysis around entry. We begin by estimating the baseline model (column 8-1) by relating *entries per contest* to *entry visibility* and including our contest fixed effects – *domain* and *site*. We find that increasing visibility (from blind to unblind) results in a significant increase in number of entries for a contest. The magnitude of this effect is over 39 additional entries per unblind tournament⁵ – a substantial 60% increase from blind cases. Because our contests occur over time, we extend the model to include *day* as an explanatory effect and *entries per day* as the dependent variable (column 8-2). We further account for dynamic contest qualities like competition by including our measures of agent experience and talent, which are all reported as cumulative daily averages for unique entrants. This means those agent measures reflect the mean experience or talent participating in the contest up to that point. The coefficient observed for *day* is positive and significant, showing that more entries arrive at the end of contests, which matches our experience with this domain and these platforms in the past. Of our experience and skill measures, *average reputation* is negative and significant – higher average skill scores of prior entrants result in fewer entries per day. Overall, these results mirror our baseline model, with significantly fewer submissions (coeff: -0.286, p-value: 0.007) in blind contests.

Increased entries in unblind contests could stem from attracting more entrants or from enticing existing agents to submit more ideas, as outlined in section 3.1. Our participant models (columns 8-3 and 8-4) address the first alternative, with negative binomial regressions using *entrants per contest* and

⁴ In assuming a negative binomial distribution for our dependent variable counts, our model includes a log link, and the resulting log-linear function can be represented as $\ln(Y_i) = \alpha + \beta(\text{Entry Visibility})_i + \delta_i + \varepsilon_i$.

⁵ Given by $\exp(4.647) - \exp(4.647 - 0.472) = 39.2$

Table 8. Comparison of Contest Productivity between Visibility Treatments

Dependent variable	8-1	8-2	8-3	8-4	8-5
	Entries per contest	Entries per day	Entrants per contest	Entrants per day	Entries per entrant
Explanatory variables	Contest fixed effects	Day, talent, experience	Contest fixed effects	Day, talent, experience	Contest fixed effects
Constant	4.647 *** (0.225)	1.708 * (0.993)	3.667 *** (0.265)	-0.213 (1.244)	0.970 *** (0.143)
Treatment					
Blind	-0.472 ** (0.185)	-0.286 *** (0.107)	-0.371 * (0.225)	-0.324 ** (0.128)	-0.078 (0.124)
Fixed effects	Yes	Yes	Yes	Yes	Yes
Explanatory Variables					
Day		0.288 *** (0.026)		0.222 *** (0.032)	
Avg. days of experience		-0.003 (0.006)		0.007 (0.007)	
Avg, contests won		-0.175 (0.114)		-0.012 (0.011)	
Avg. reputation (z-score)		-0.722 *** (0.245)		-0.575 * (0.322)	
Chi-squared test	9.6	102.2	5.1	76.6	14.7
Mean Response	83.10	12.09	28.13	5.98	2.96
Observations	8	55	8	55	225
DF	5	9	5	9	5

Negative binomial regression on contest productivity counts, base case is Unblind visibility

Significance levels: * <0.10, ** <0.05, *** < 0.01

Standard errors given in parentheses

Note: 8-4 reports on every unique daily entrant; robust to excluding entrants who submitted on prior days.

Mean Response listed is log-transformed to show actual values for dependent variable measure.

entrants per day as the dependent variables. In both cases, more agents are choosing to participate in unblind contests. The magnitude of this effect is about 12 more entrants per unblind contest.⁶ It should be noted that our baseline entrant model (column 8-3) isn't significant at the overall level, even though its message is consistent. As before, our daily model (column 8-4) presents a better characterization of the contest environment and includes controls for dynamic contest elements – both time and participant characteristics.⁷ The coefficient for entrants is negative and significant (coeff: -0.324, p-value: 0.011), suggesting that fewer entrants participate in blind contests. If we look specifically at *entries per entrant* across the contests (column 8-5), we see no differences in behavior, with 225 agents submitting on average 2.96 entries per contest regardless of entry visibility. This resonates nicely with the prior finding that submitting extra entries in unblind contests doesn't increase an agent's chance of winning (Bockstedt et al. 2011).

⁶ Given by $\exp(3.667) - \exp(3.667 - 0.371) = 12.1$

⁷ We exclude participant characteristic variables from 8-1 and 8-3 because of limited degrees of freedom.

These results support Hypothesis 1. Unblind contests generate more entries; however they do this *not* by inducing more entries from existing agents but by increasing the number of agents that participate. An interpretation of this result is that entry visibility reduces barriers to entry, allowing easier exploration of the search landscape and enticing more agents to search for a solution and submit.

We also perform a number of robustness checks to address potential concerns. In addition to *day* (our control for how much of a contest has elapsed), we add a categorical *day of the week* variable to account for the fact that some days might have different behavior patterns (i.e., Saturday may see fewer agents online). Another concern might be that the dependent variables in our daily models (8-2 and 8-4) might suffer from serial correlation across days. In addition to several tests (such as Durbin-Watson calculations via STATA and by hand) that indicate no auto-correlation, we add lagged variables for entries and entrants to our specification and observe results consistent with our daily models. Finally, we address alternate calculations of experience and talent, including (a) totals instead of averages and (b) rolling averages that are inclusive of repeat entrants instead of inclusive. In each of these cases, we find no differences in our primary findings.

To corroborate these results and further check robustness, we run a similar analysis on 44,582 logo contests from 99Designs from 2009-2012. These contests were those conducted in \$US and unlocked (meaning a winner was chosen and the site rules followed), and they account for 78% of all logo contests during that time. We include fixed effects to control for contest characteristics that were stable in our experiments, including prize amount, prize guarantee, and level of engagement from the administrator. The results (in Appendix B) show that blind contests enjoy significantly fewer entries (coeff: -0.33^{***}) and entrants (coeff: -0.29^{***}), similar to our experiment results in Table 8. The size of this visibility effect in the panel regression is roughly equivalent to \$83 in prize money, so making a contest's entries blind has the same participation effect as reducing the prize by 23%. These findings add additional evidence that our experiment results are both directionally correct and consistent with broader contest trends for these platforms.

7.2 Characteristics of Entries – Similarity

Having shown that the entry decision varies with entry visibility, we now turn to how that behavior impacts the search process. We use our pairwise similarity measures (from section 4.2) to determine whether designers create submissions that are more similar when they are permitted to see others' entries. First, we examine the contests at an aggregate level by comparing the *average contest similarity* with a simple t-test on the means (Table 9). Average contest similarity is the mean of all possible pairwise similarities within a contest. This captures, independent of when logos were submitted, how alike our lab group believed a contest's entries to be. We find that in aggregate, average similarity in unblind contests is approximately 14% greater and significant (0.194 vs. 0.170; t-statistic 2.59). This is meaningful, and supports Hypothesis 2, but to better capture the degree to which agents are incorporating elements from prior designs, we extend our analysis.

Table 9. Comparison of Idea Similarity within Contests

Overall contest similarity score (mean of all contest pairwise scores):

	Blind	Unblind
Average Pairwise Similarity	0.170	0.193
Number of observations	4	4

T-statistic: 2.59 **

Significance levels: * <0.10, ** <0.05, *** < 0.01 (two-tailed test)

Overall contest similarity score is mean of each pairwise score in a contest.

Robust to more conservative measures (i.e., omitting any pairs from the same designer)

Logo-level similarity scores:

	9-1	9-2	9-3
Dependent variable	Similarity to prior entries	Similarity to prior entries	Similarity to prior entries
Explanatory variables	Contest fixed effects	Interaction with day	Interaction with period
Constant	0.230 *** (0.016)	0.406 *** (0.027)	0.724 *** (0.037)
Treatment			
Blind	-0.038 *** (0.014)	-0.158 *** (0.039)	-0.529 *** (0.058)
Fixed effects	Yes	Yes	Yes
Timing			
Day		-0.034 *** (0.004)	
Day x Blind		0.024 *** (0.007)	
Period			-0.509 *** (0.035)
Period x Blind			0.506 *** (0.058)
R-squared	0.03	0.12	0.27
Mean Response	0.21	0.21	0.21
Observations	633	633	633
DF	5	7	7

OLS regression on idea similarity scores, base case is Unblind visibility

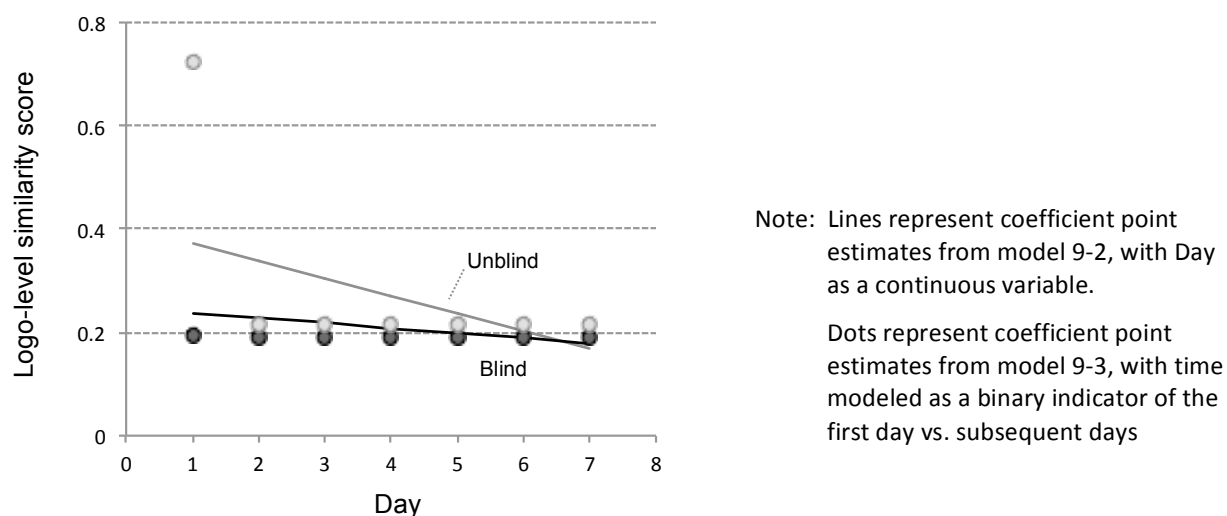
Significance levels: * <0.10, ** <0.05, *** < 0.01

Standard errors given in parentheses

Note: Logo-level similarity score is mean similarity of each entry to prior entries. Period is binary and defined as the first day (0) or days two through seven (1). Robust to only first entries.

Model 9-1 shows the baseline results of our linear regression for submission similarity. Our dependent variable is *logo-level similarity*, which for each entry is the degree of similarity to *prior* submissions. We find that increasing visibility (from blind to unblind) results in entries that are significantly more similar. The magnitude of the effect is such that unblind contests were rated as 20% more similar. Including time effects (column 9-2), however, notable differences emerge. While entries in unblind contests are much more similar initially, by the final day, that difference has been erased. At that point, entries to unblind contests are just as unique as those in their blind counterparts. Probing a bit further (column 9-3), we can use period categorical variables to see that the difference between blind and unblind contests in terms of entry similarity happens almost exclusively in the first day of our experiments. After that, there is no discernible difference between the treatments. Figure 3 highlights these composite effects, with the coefficient point estimates for columns 9-2 and 9-3 represented as lines and points, respectively.

Figure 3. Entry Similarity over Time



The implication is that while unblind contests do encourage submissions that are more similar, that phenomenon is limited to the early stages of the contest. Several things could be happening. This could be because participants only incorporate elements from prior entries early in the process. More likely, once there is a sufficient breadth of entries, inspiration will have more seeds from which to spring and the resulting conformity will be harder to detect. This could be the result of a diffusion process, in which an initial seed is planted and ideas radiate out from that seed. As the ideas radiate out into a larger area, there are a greater number of seeds from which to create an incremental variant and average similarity declines. This explanation is plausible, given the results over the course of the contest. The data suggests that by increasing the visibility within tournaments, resulting submission are more similar, but that this effect quickly disappears. On balance, it appears to not overwhelm the pool of entries with conformity, which is

beneficial from the administrator's standpoint.

7.3 Characteristics of Entries – Quality

Table 10 shows the results of a regression analysis with logo *rating* as the dependent variable; our explanatory variables and contest fixed effects (section 4.3) are also included. We use a clustered OLS because there are multiple ratings for each logo and our explanatory variables are observed at the level of the logo, not the level of the rating. In our baseline model (column 10-1), we find that blind contests result in higher quality entries. This result is marginally significant and in the *opposite* direction of our hypothesis, which predicted that unblind contests would return better entries on average. Recall that there was some evidence of such a relationship, but we believed the balance of evidence would push the net

Table 10. Comparison of Contest Quality between Visibility Treatments

Dependent variable	10-1 Ratings (all entries)	10-2 Ratings (all entries)	10-3 Ratings (re-submits)
Explanatory variables	Number of entries	Contest results	Contest results
Constant	2.098 *** (0.111)	2.010 *** (0.130)	0.493 ** (0.225)
Treatment			
Blind	0.084 * (0.051)	-0.232 * (0.131)	-0.366 (0.254)
Fixed effects	Yes	Yes	Yes
Explanatory variables			
Entries	0.001 (0.001)	0.001 (0.001)	-0.001 (0.001)
Entries x Blind		0.003 (0.003)	0.007 *** (0.002)
Max		0.070 *** (0.023)	0.114 *** (0.027)
Max x Blind		0.009 (0.043)	-0.124 *** (0.045)
Personal best		0.023 (0.021)	0.528 *** (0.068)
Personal best x Blind		0.094 *** (0.033)	0.243 *** (0.085)
Reputation		0.154 *** (0.027)	0.033 (0.035)
Reputation x Blind		0.058 (0.047)	-0.006 (0.049)
R-squared	0.08	0.11	0.20
Mean Response	2.41	2.41	2.40
Observations	11,380	11,380	7,050
DF	45	52	52

OLS regression on individual ratings, clustered by logo, base case is Unblind visibility

Significance levels: * <0.10, ** <0.05, *** < 0.01

Robust clustered standard errors given in parentheses

Note: Robust to different measures of agent expertise, including an agent's highest/final *personal best*.

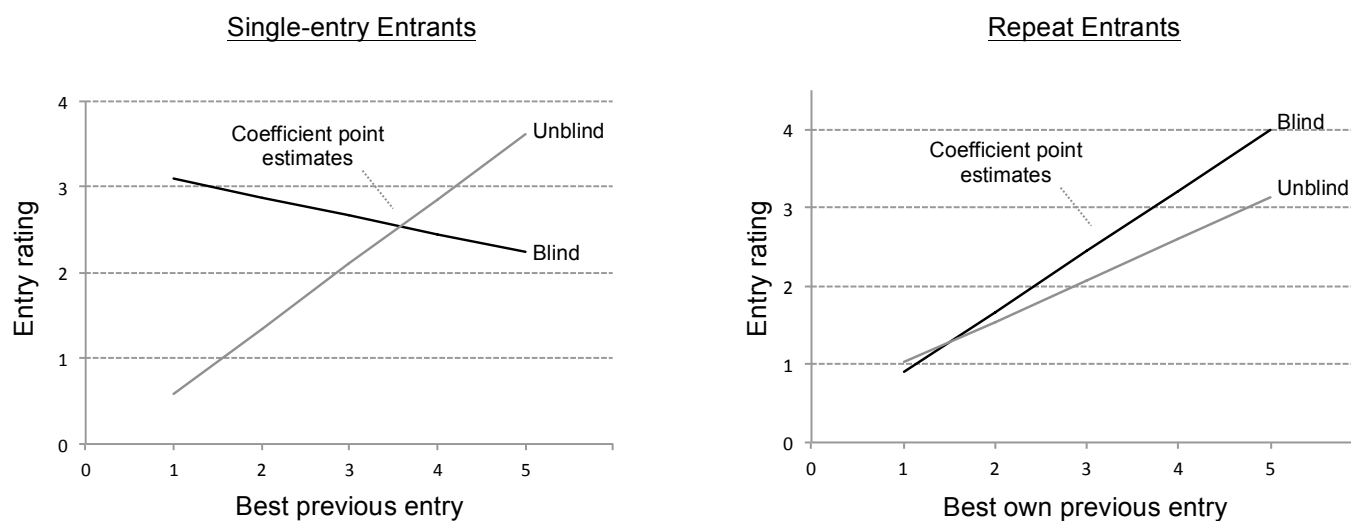
effect in the other direction. Agent talent was a key determinant in that argument, so we attempt to approximate agent expertise and explore this result further.

In addition to our platform measure of skill (*reputation*), we also include *personal best* – a variable that captures the best score an individual has received on prior submissions. If success were random and previous scores weren't predictive of future entries for a given agent, then this metric would be ineffective. However, if we ignore agents who only submit once (for which there are no prior scores), the correlation between *personal best* and rating is 0.30. If we look at an agent's highest *personal best* globally and compare that talent measure to all their ratings, the correlation is 0.71. In a noisy environment, it indicates there is information in this measure of performance.

Interestingly, when we include our explanatory variables to control for the amount of information in the contest and the performance of the designers, our main effect switches signs (column 10-2). Now, we observe that the previous result of blind being better (column 10-1) seems to be partially driven by better designers opting in to a few of the contests. A simple t-test on entrant reputation shows no difference in blind versus unblind contests overall. The effect of visibility differs, however, based on agent talent. We observe that unblind contests are better for new entrants (who have no previous best entry) and low-quality designers. High-quality agents perform better in blind contests. Thus, the benefit of entry visibility depends on the type of participants in a given contest.

For this reason, we test one further extension by explicitly modeling just repeat submitters (column 10-3). In this case, low-quality designers in blind and unblind contests submit entries that are identical in quality. As expertise grows, submission quality improves more for blind contests, mirroring the result in column 10-2. If repeating agents are strictly better off in the blind condition, then it is one-time entrants who benefit disproportionately from entry visibility (Figure 4). This lends additional strength to our theory that unblind contests add value by lowering the barriers to entry. Those low-effort designers, who don't submit more than once, benefit from being able to see high quality entries. Looking at *Max* – the best prior entry by others – we see that with submission visibility, new entries mirror the best existing quality and appear anchored to past results. This effect completely goes away in the blind case, as one would expect. If agents can't see other entries, submission quality decreases with better prior entries, consistent with economic theory around incentive effects (Boudreau et al. 2011). These findings support our hypothesis in part, but also add a new layer of understanding to the tournament literature.

Our final measure of interest is variance in quality. Table 11 starts with a baseline model (column 8-1) that relates our de-trended measure of *quality variance* to *entry visibility* and includes contest fixed effects – *domain* and *site*. We find that increasing visibility (from blind to unblind) reduces the variance in quality we see in the submission ratings. When including *day* and *number of entries* as explanatory variables (columns 8-2 and 8-3), our results hold, with variation in the blind setting growing with number of entries. This trend is reasonable; the differences in contest visibility grow over time, as more aggregate information is available. These findings are consistent with Hypothesis 4.

Figure 4. Submission Quality given Search Landscape

Note: Repeat Entrants graph given by coefficient point estimates from model 10-3;
Single-entry Entrants from modified model 10-2 (with only single-entry entrants)

Table 11. Comparison of Contest Variance between Visibility Treatments

Dependent variable	11-1	11-2	11-3
	Variance (de-trended)	Variance (de-trended)	Variance (de-trended)
Explanatory variables	Contest fixed effects	Contest day	Interaction with entries
Constant	0.464 *** (0.034)	0.408 *** (0.047)	0.454 *** (0.049)
Treatment			
Blind	0.090 *** (0.028)	0.088 *** (0.028)	0.023 (0.047)
Fixed effects	Yes	Yes	Yes
Timing			
Day		0.011 * (0.007)	-0.014 (0.011)
Entries			
Entries			0.001 ** (0.001)
Entries x Blind			0.003 ** (0.001)
R-squared	0.13	0.13	0.15
Mean Response	0.26	0.26	0.26
Observations	665	665	665
DF	5	6	8

OLS regression on quality de-trended variation at the logo level, adjusted for Judge, base case is Unblind
Significance levels: * <0.10, ** <0.05, *** < 0.01
Standard errors given in parentheses

8 Discussion

To understand and characterize the implications of a relatively new decision afforded to innovation contest administrators – that of entry visibility – we examined two primary pathways of influence: (1) the likelihood of entry from an agent and (2) the resulting characteristics of entries in a contest. The related hypotheses we pose in Section 3 are largely supported.

In addressing the first pathway, we find that unblind contests generate more entries; however they do this *not* by inducing more entries from existing agents but by increasing the number of agents that participate. For the second pathway, we examine characteristics for both submission similarity and the quality distribution of entries and find the effect of visibility depends on the setting. Unblind contests encourage submissions that are more similar, mostly in the early stages of the contest. For single-entry participants, entry quality “ratchets up” with the best previous entry if it’s visible, while moving in the opposite direction if it’s not. However, for invested participants who submit more than once, those with better prior submissions improve more in the absence of entry visibility. Variance in entry quality also improves in the absence of entry visibility.

8.1 Managerial Implications

This research is motivated by the managerial question of whether or not an innovation tournament administrator can improve outcomes based on the moderating decisions within the contest. We found strong evidence to suggest that there are very real differences that result from those decisions. While we cannot extrapolate our results to all innovation contests, understanding the implications of participant entry, idea similarity from search, and contest outcomes should permit managers to more effectively tailor contests for optimal output. Specifically, we uncovered three key decisions contest administrators should manage.

First, managers should be aware that barriers to entry are an important consideration. Unblind contests can attract more entrants, likely because they permit easier search. Casual observers can see exemplars to kick-start their idea development. This doesn’t increase the number of entries submitted by each solver, but it does get more solvers in the door.

Second, participant motivation has an effect. The learning environment of unblind contests is better than in blind contests for participants that only submit one entry; seeing a good entry prompts them to come up with a better submission. This is not the case for repeat submitters, who produce better ideas in blind contests. So in an internal company tournament where employees are motivated to participate and likely to submit multiple entries, blind contests may promote better quality (and more varied) ideas. However, in a crowdsourced contest via social media, unblind contests will likely provide better access to landscape exploration and learning and consequently a better result.

Third, entry visibility does impact similarity in designs, but less than we imagined. Unblind contests see a higher level of similarity than blind contests, but the effect quickly goes away. The fear that

designers will get stuck in one part of the search landscape does not manifest itself in our data. It appears that participants can use other submissions and create incremental variants that are sufficiently different quickly and efficiently.

8.2 Limitations

Given the fact that we performed this experiment with real designers instead of in a lab, we could not use some potentially interesting designs. The reaction of the same individual under different treatments would be interesting and potentially feasible in a lab study, although in our case, it was not possible.

While the contests were constructed to be nearly identical, we made slight changes in the details in order to avoid detection in the marketplaces. Although we control for contest fixed effects, different challenges could attract fundamentally different types of agents, which could introduce unaccounted for bias into our model.

Beyond this, the backdrop that served for our study deserves some mention. Our setting is nice in that it uses real marketplaces and real designers to test these theories. In addition, logos and graphic design are nice in that the whole idea is represented visually. This may help give insight into more complex domains. However, in graphic design contests (such as those around logos), the effort needed to produce any single idea is relatively small, which could also have implications. Unblind contests may be more acceptable in such situations because the level of investment is minimal. Contests requiring more substantial investment or areas with substantial benefits to intellectual property may not flourish under the same conditions.

8.3 Future Work

As the first to look at the differences of entry visibility on innovation contest outcomes, we have just begun our understanding of this moderating decision. The following questions seem promising for future exploration:

- How does entry visibility apply to different settings? There are plenty of administrator decisions that could improve performance depending on the characteristics of the contest, the solvers, and their interaction with the entry visibility design choice.
- Do different classes of problems behave in the same way? Do algorithmic contests match graphic design contests as related to entry visibility?
- If similarity between ideas does get lost in the unblind case fairly quickly, what density of solvers or entries would be required to again pick up on similarity in ideas? Would a less densely populated ideation landscape change this finding?
- Diverse perspectives are seen as a benefit of open innovation (Jeppesen and Lakhani 2010). Here, we used a pool of solvers from an established contest platform. Controlling for innate solver characteristics would be an interesting direction to further extend the understanding from the level of the agent.

References

- ALICKE, M. D., M. L. KLOTZ, D. L. BREITENBECHER, T. J. YURAK, and D. S. VREDENBURG (1995): "Personal Contact, Individuation, and the Better-than-Average Effect," *Journal of Personality and Social Psychology*, 68, 804-825.
- ÅSTEBRO, T., S. A. JEFFREY, and G. K. ADOMDZA (2007): "Inventor Perseverance After Being Told to Quit: The Role of Cognitive Biases." *Journal of Behavioral Decision Making*, 20, 253-272.
- BOCKSTEDT, J., A. MISHRA, and C. DRUEHL (2012): "Do Participation Strategy and Experience Impact the Likelihood of Winning in Unblind Innovation Contests?," working paper.
- BOUDREAU, K. J., N. LACETERA, and K. R. LAKHANI (2011): "Incentives and Problem Uncertainty in Innovation Contests: An Empirical Analysis," *Management Science*, 57, 843-863.
- BULLINGER, A. C., A. NEYER, M. RASS, and K. M. MOESLEIN (2010): "Community-Based Innovation Contests: Where Competition Meets Cooperation," *Creativity and Innovation Management*, 19, 290-303.
- CASAS-ARCE, P., and F. A. MARTÍNEZ-JEREZ (2009): "Relative Performance Compensation, Contests, and Dynamic Incentives," *Management Science*, 55, 1306-1320.
- CHE, Y. and I. GALE (2003): "Optimal Design of Research Contests," *The American Economic Review*, 93, 646-671.
- DAHAN, E. and H. MENDELSON (2001): "An Extreme Value Model of Concept Testing," *Management Science*, 47, 102-116.
- DAHL, D. W. and P. MOREAU (2002): "The Influence and Value of Analogical Thinking During New Product Ideation," *Journal of Marketing Research*, 39, 47-60.
- DOW, S. P., J. FORTUNA, D. SCHWARTZ, B. ALTRINGER, D. L. SCHWARTZ, and S. R. KLEMMER (2012): "Prototyping Dynamics: Sharing Multiple Designs Improves Exploration, Group Rapport, and Results," *Design Thinking Research*, 47-70.
- ERAT, S. and V. KRISHNAN (2012): "Managing Delegated Search Over Design Spaces," *Management Science*, 58, 606-623.
- FULLERTON, R. L. and R. P. McAFEE (1999): "Auctioning Entry into Tournaments," *American Economic Review*, 107, 573-605.
- GIROTRA, K., C. TERWIESCH, and K. T. ULRICH (2010): "Idea Generation and the Quality of the Best Idea," *Management Science*, 56, 591-605.
- GUO, S. and M. W. FRASER (2009): *Propensity Score Analysis: Statistical Methods and Applications*, Sage Publications.
- HAMILTON, B. H. and J. A. Nickerson (2003). "Correcting for Endogeneity in Strategic Management Research," *Strategic Organization*, 1, 51-78.
- HECKMAN, J. (1979): "Sample selection bias as a specification error," *Econometrica*, 47, 153-61.
- HILBE, J. M. (2011): *Negative Binomial Regression*, Cambridge University Press.
- JEPPESEN, L. B. and K. R. LAKHANI (2010): "Marginality and Problem-Solving Effectiveness in Broadcast Search," *Organization Science*, 21, 1016-1033.
- KAHNEMAN, D. and A. TVERSKY (1977): *Intuitive Prediction: Biases and Corrective Procedures*, McLean VA: Decisions and Designs Inc.

- KORNISH, L. and K. ULRICH (2011): "Opportunity Spaces in Innovation: Empirical Analysis of Large Samples of Ideas," *Management Science*, 57, 107-128.
- LAKHANI, K. R., L. B. JEPPESEN, P. A. LOHSE, and J. A. PANETTA (2007): "The Value of Openness in Scientific Problem Solving," Harvard Business School Working Paper, 07-050.
- LAZEAR, E. P., and S. ROSEN (1981): "Rank-order Tournaments as Optimum Labor Contracts," *Journal of Political Economics*, 89, 841-864.
- LEE, B., S. SRIVASTAVA, R. KUMAR, R. BRAFMAN, and S. R. KLEMMER (2010): "Designing with Interactive Example Galleries," Proceedings of the conference on Human Factors in Computing Systems, ACM, 2257-2266.
- MARCH, J. (1991): "Exploration and Exploitation in Organizational Learning," *Organization Science*, 2, 71-87.
- MARSH, R. L., J. D. LANDAU, and J. L. HICKS (1996): "How Examples May (and May Not) Constrain Creativity," *Memory & Cognition*, 24, 669-680.
- MERTON, R. K. (1942): *The Sociology of Science: Theoretical and Empirical Investigation*, Chicago: University of Chicago Press.
- METCALFE, J.S. (1994): "The Economics of Evolution and the Economics of Technology Policy," *Economic Journal*, 104, 931-944.
- MULKAY, M. (1975): "Three Models of Scientific Development," *Sociological Review*, 23, 509-526.
- NELSON, R. R. and S. WINTER (1982): *An Evolutionary Theory of Economic Change*, Cambridge, MA: Harvard University Press.
- SMITH, S. M., T. B. WARD, and J. S. SCHUMACHER (1993): "Constraining Effects of Examples in a Creative Generation Task," *Memory & Cognition*, 21, 837-845.
- TABACHNICK, B. G and L. S. FIDELL (2007). *Using Multivariate Statistics*, Boston: Allyn and Bacon.
- TAYLOR, C. R. (1995): "Digging for Golden Carrots: An Analysis of Research Tournaments," *American Economic Review*, 85, 872-890.
- TERWIESCH, C. and K. T. ULRICH (2009): *Innovation Tournaments: Creating and Selecting Exceptional Opportunities*, Harvard Business School Press.
- TERWIESCH, C. and Y. XU (2008): "Innovation Contests, Open Innovation, and Multiagent Problem Solving," *Management Science*, 54, 1529-1543.
- VON HIPPEL, E. (2005): *Democratizing Innovation*, Cambridge, MA: MIT Press.
- WALTER, T. and A. BACK (2011): "Towards Measuring Crowdsourcing Success: An Empirical Study on Effects of External Factors in Online Idea Contest," 6th Mediterranean Conference on Information Systems.
- WARD, T. B. (1994): "Structured Imagination: The Role of Category Structure in Exemplar Generation," *Cognitive Psychology*, 27, 1-40.
- WOOTEN, J., and K. ULRICH (2014): "Idea Generation and the Role of Feedback: Evidence from Field Experiments with Innovation Tournaments," working paper.
- YANG, Y., P. CHEN, and P. PAVLOU (2010): "Managing Open Innovation Contests in Online Market," working paper.

Appendix A. Sample Design Briefs

99Designs – Power Perk Coffee

BUSINESS NAME:

Power Perk Coffee

DESCRIPTION OF BUSINESS:

A better cup of coffee at home – Power Perk focuses on the best ingredients, processes, and accessories for coffee drinkers.

PREFERRED LOGO TYPES:

None specified

COLOR PREFERENCES:

No restrictions on color

TO BE USED ON:

Print (Business cards, letterheads, brochures etc.)
 Online (Website, online advertising, banner ads etc.)
 Merchandise (Mugs, T-shirts etc.)

NOTES:

Branding - Logo should work across the entire line of coffee products (beans, percolators, and accessories).

Demographics - Our target audience is young adult coffee drinkers (18-35 years old) in the US who are college-educated.

CrowdSpring – Bold Brew Tea

WHAT IS THE EXACT NAME YOU WOULD LIKE IN YOUR LOGO?

Bold Brew Tea

DO YOU HAVE ANY OTHER INFO OR LINKS YOU WANT TO SHARE?

Industry - Home Tea Brewing. Tea leaves, brewing systems, and other accessories for tea drinkers.

Demographics - The focus is on the young adult market in the US. 18-35 year olds who are college-educated and discovering tea as a great beverage alternative.

WHAT ARE THE TOP 3 THINGS YOU'D LIKE TO COMMUNICATE THROUGH YOUR LOGO?

The brand should work over the whole line of tea products. High quality ingredients and processes are the foundation for our image and great-tasting product.

WHAT LOGO STYLES DO YOU LIKE (IMAGE + TEXT, IMAGE ONLY, TEXT ONLY, ETC.)

- Any colors/styles
- No restrictions

Appendix B. Comparison of Contest Productivity and Visibility on 99Designs (whole site)

Dependent variable	B-1	B-2
	Entries per contest	Entrants per contest
Constant	2.943 *** (0.015)	2.121 *** (0.014)
Explanatory variables		
Blind	-0.333 *** (0.010)	-0.289 *** (0.009)
Prize amount (\$US)	0.003 *** (0.000)	0.003 *** (0.000)
Guaranteed prize	0.468 *** (0.006)	0.452 *** (0.006)
Engagement (% scored)	0.806 *** (0.015)	0.165 *** (0.015)
Chi-squared test	28,388.0	27,770.7
Mean Response	132.95	34.16
Observations	44,582	44,582
DF	4	4

Negative binomial regression on contest productivity counts

Significance levels: * <0.10, ** <0.05, *** < 0.01

Standard errors given in parentheses

Notes: Mean Response listed is log-transformed to show actual values for dependent variable measure.

Blind results in fewer entries and entrants in 3rd party contests, corroborating our experiments.

Bigger % impact on entrants than entries, also corroborating. Blind same as reducing prize by \$83.

Data: 44,582 logo contests (2009-12; in US\$ with unlocked archives; 78% of all logo contests)