Innovation Field Experiments: Empirical Insights and Policy Implications from Organizing Innovation Contests

Karim R. Lakhani | Harvard University
Kevin J. Boudreau | London Business School

Draft April 6, 2015

Abstract: In this paper we discuss how field experiments can be applied to research and policy questions concerned with the design and organization of innovation contests. Although the theory of innovation contests is well advanced, there is a dearth of empirical evidence on how theory reflects observed behavior and its implications for their general deployment as a routine mechanism to elicit innovation. Results from an ongoing field experimental program at the Crowd Innovation Laboratory at Harvard University are presented to illustrate how causal explanations regarding the role of incentives, knowledge and search process can be derived for the innovation literature. The research program has simultaneously solved important innovation challenges for partner organizations like NASA and Harvard Medical School while simultaneously contributing experimental evidence of interest to the innovation literature.

Innovation Field Experiments: Empirical Insights and Policy Implications from Organizing Innovation Contests

I. Introduction

In empirical economics, the most common research approach entails combining theoretical perspectives with an application of appropriate econometric techniques to naturally occurring data in order to obtain causal explanations (List 2009). In essence, empirical economists, regardless of their chosen data sources and estimation approaches, when reporting results are documenting the outcomes of an "experiment" (Harrison and List 2004). Researchers are typically interested in completely isolating the effect of a "treatment" on a population of interest, and the various empirical approaches aim to generate causal claims, with varied degrees of success in achieving this objective. Although prior paradigmatic views about the potential of economists to run controlled experiments were quite negative and relegated economic science to natural data obtained from observational methodologies similar to those pursued by astronomers and meteorologists (Samuelson and Nordhaus 1985), over the last two decades there has been a boom in the discipline with a rapid embrace of laboratory and field experimental approaches (List 2011).

These experimental approaches provide researchers precise control over the data generation process and enable the application of exogenous treatments of interest to a randomized sample of representative participants, approximating the methods available to natural scientists such as chemists, biologists, physicians and medical scientists and physicists. Although laboratory-based economic experiments now have established experimental infrastructure available at most research institutions, embracing the field experimental methodological paradigm requires significant commitment and investment by the researcher (Duflo and Banerjee 2009, Bandiera et al. 2011). In addition to developing a theory-based perspective on an appropriate research question, scholars engaged in field experiments need to acquire substantial resources to gain cooperation from the field setting (e.g. an organization that provides access), design exogenous treatments, obtain institutional review board approval, select the tasks or actions to be studied, recruit representative subjects, ensure treatment and control isolation, record data, provide results and feedback to the sponsor and then proceed with data analysis.

This places a high barrier for researchers interested in pursuing field experiments. Despite these barriers, field experimental approaches have appeared to gain substantial ground in a variety of sub-disciplines including behavioral, developmental and labor economics (Bandiera et al. 2011, List 2009, 2011).

The economics of innovation literature has generally lagged behind in adopting the field experimental approach. Beyond the obstacles noted above, innovation researchers face the challenge that conducting an experiment requires deep cooperation and collaboration with organizations engaged in research and development activities. Naturally, many executives would be reluctant to allow research access to a setting that is responsible for generating innovation, with significant concerns that organizational goals are not compromised due to the application of randomized treatment.

In this paper we review a systematic field experimental research program undertaken by the Crowd Innovation Laboratory (CIL) at Harvard University's Institute for Quantitative Social Science. This program represents an early first step towards bringing the field experimental research method to the economics of innovation literature. The laboratory focuses its efforts on the design of contests that create solutions to real-life technological problems and questions regarding the optimal organization of scientific activity.

The impetus for a systematic program to develop empirical insights on the organization of innovation contests has arrived from policy makers and scholars. A report investigating the feasibility of implementing innovation contests[1] at the National Science Foundation (NSF) by the National Research Council (2007: page 11) highlighted that "owing to the limited experience with innovation prizes, relatively little is known about how they work in practice or how effective they may be as compared with, for example, R&D grants and contracts, or tax incentives." Beyond comparing the relative effectiveness of the various institutions for encouraging innovation, several scholars have also noted that there is a paucity of empirical evidence, as compared to the advanced stage of economic theories, on the role of prizes and contests to induce

[1] We use prizes, tournaments and contests interchangeably to denote institutions for innovation that provide performance-contingent incentives.

innovation (e.g.: Brunt et al. 2011, Murray et al. 2012, Williams 2012, Boudreau et al. 2011). Hence policy makers and researchers interested in innovation prizes need to both understand if contests deliver superior performance as compared to alternative mechanisms and how behavior and actions in contests match or depart predictions in theory.

The CIL aims to develop empirical insights on the operation, effectiveness and related theory underlying innovation contests. The emergence of several online platforms (e.g. InnoCentive, Kaggle, Quirky, TopCoder, Threadless) that harness hundreds of thousands of problem solvers to engage in a variety of innovation-related problem solving activities (Boudreau and Lakhani 2013) provides a unique opportunity to apply the field experimental approach to relevant innovation tasks and innovators. Beyond the online setting, the laboratory has assisted in the running of several scientific grant application processes, essentially an internal contest amongst researchers to win funding, while simultaneously layering in appropriate randomization to derive causal economic insights.

The CIL, established in 2010, designs and executes contests in close partnership with the U.S. National Aeronautics and Aerospace Administration (NASA), Harvard Medical School (HMS) and TopCoder (an online innovation contest platform) to help solve their innovation problems and develop evidence-based advice on the organization of the innovation process. Through the design of field experiments, the research program also simultaneously investigates causal explanations for the behaviors of participants in various types of contest settings and activities.

A recent innovation contest using a problem in computational biology illustrates how field experiments can create concrete working solutions, provide comparative evidence for performance and create insights on underlying mechanisms of interest to innovation scholars. Colleagues at HMS approached the CIL with an interest in understanding the applicability of having “crowds,” in the form of large numbers of loosely affiliated individuals, engage in solution generation for problems originating from a large-scale academic medical center. Working closely with the scientists at the Harvard Catalyst, Harvard's Clinical and Translational Science Center, we identified a challenging computational biology problem in immunogenomics (the genetic study of the immune system) that could serve as a basis for a contest and developed an automated scoring system to evaluate contest submissions. In cooperation with TopCoder, we posted the problem to their membership and, over the course of two weeks, 122 solvers (out of the 722 signups) from 89 countries created 650 solutions for a total prize purse of $6,000. Our analysis (Lakhani et al. 2015) revealed that 30 solutions exceeded the performance of the internal Harvard solution and the benchmark National Institutes of Health (NIH) developed approach, and the best of them advanced the state of the art by a factor of 1,000. Also surprising was the short duration of time required to create the solutions (just two weeks), as compared to years of grant-supported internal academic efforts, and the fact that none of the solvers had any background in computational biology.

The contest not only solved a computational bottleneck in immunogenomics, it also yielded empirical insights on the mechanisms underlying how crowds are utilized for innovation. Boudreau and Lakhani (2009, 2013) and King and Lakhani (2013) have clarified that participants in crowds can be organized in the form of contests, where they are competing to solve a problem, or managed via communities (e.g. open source software projects (Lerner and Tirole 2002)) in which a collective solution is developed. Layered within the solution development process was a field experiment that randomized knowledge disclosure treatments, which simulated conditions of intermediate disclosure observed in a community setting, where all participants have access to the work of others during the problem solving period, and final disclosure in a contest, where solution approaches are revealed upon awarding of the prize (Boudreau and Lakhani 2015). Indeed, the timing and form of knowledge disclosures are fundamental properties of many of society's innovation systems (Scotchmer 2002).

Examination of the randomized treatments revealed that intermediate disclosure depressed incentives, resulting in lower participation and effort yet achieving higher innovative performance as compared to participants working under final disclosure. A close study of the software submissions allowed us to show that the performance differences could be accounted for through the qualitatively different search processes induced by the treatments. Intermediate disclosure (community) subjects, with access to the work of others, converged their search to a few technological pathways that demonstrated promising early results, while final disclosure (contest) subjects, working independently, generated more diverse and novel solution approaches that tended to be of lower quality. The field experimental design enabled a *ceteris paribus* evaluation of the impact of knowledge disclosure on innovation that is difficult to achieve using observational data, as the institutions of contests and communities typically operate on problems and settings that are not similar.

The above exercise provides a prototypical view of how field experiments can combine salient questions from the economics of innovation literature while simultaneously solving a relevant innovation problem for the host organization. The remainder of the paper proceeds as follows. Section II provides the rationale behind the establishment of the CIL and its operational approach in working with sponsors. We also highlight the range of innovation problems that have been brought through the laboratory. In Section III, we discuss the framework that guides the design of the field experiments and provide results from four studies. We then outline three field experimental design parameters as they relate to innovation activities in Section IV. Section V concludes.

II Rationale and Operational Approach of the Crowd Innovation Laboratory

The mission of the Crowd Innovation Laboratory (CIL) at the Harvard Institute for Quantitative Social Science is to simultaneously solve our partners' innovation challenges while pursuing core social science questions through the implementation of randomized controlled field experiments on topics related to innovation contests. Although the theory on contests is relatively well-advanced, empirical evidence has been very difficult to amass and has primarily relied on sports data. The CIL's work has begun to rectify this gap and has taken canonical theories from the textbook to the field. Prior to the establishment of the laboratory, we had completed research on understanding the factors underlying performance on innovation contest platforms like InnoCentive (Jeppesen and Lakhani 2010) and TopCoder (Boudreau, Lacetera and Lakhani 2011). Exposure to the platforms, detailed knowledge about their operations, and the openness of the executives to pursue further studies laid the groundwork for the possibility of conducting field experiments.

The impetus to establish the laboratory, however, came from interactions with NASA and HMS personnel, separately, in the executive education classroom. Executives from both organizations were intrigued by the performance results demonstrated in our analysis of the naturally occurring data from both platforms. Both organizations requested assistance in developing pilot programs to assess how external innovation contests could be deployed for their own internal innovation challenges. We recognized this as an opportunity to use these pilots to explore how we could accomplish the natural science and social science mission simultaneously. TopCoder executives were willing to let us modify their contest platform to suit our experimental needs, the host organizations helped us source appropriate computational problems and we raised the funds to generate the cash prizes (through generous research grants from Harvard Business School and London Business School). Both pilots vastly exceeded the expectations of the sponsor organizations in terms of the innovation results achieved and yielded the first ever, to our knowledge, field experiments in the economics of innovation literature (Section III summarizes the pilot research results).

NASA executives in particular were interested in further assessing and investigating if external contests could provide a cost effective means of generating high quality solutions to a range of computational problems. NASA released a request for proposals for an organization that would assist the space agency with identifying problems that could be solved through contests, designing and executing the contests, developing comparative cost assessments and furthering the science behind the economics of innovation contests. A joint proposal between Harvard University and TopCoder was successful in winning the contract and thus the laboratory was established. [2]

Since its establishment in 2010, the CIL has helped NASA run software innovation contests in domains as varied as asteroid detection, astronaut health applications, space station solar array positioning, planetary data evaluation, deep space disruption tolerant

[2] Initially we were called the NASA Tournament Lab to reflect our focus on contests for NASA. Today our partners have expanded beyond NASA and hence the new name.

networking and space robotics. The passage of the America COMPETES Act in 2012 provided incentive-based prize procurement authority to all federal government agencies, resulting in the White House requesting NASA to assist other federal agencies with their innovation contest projects.

The CIL has worked closely with the Center of Excellence for Collaborative Innovation (CoECI) and assisted agencies in designing innovation contests as varied as the Center for Medicare and Medicaid Services (CMS), the Environmental Protection Agency, the Department of Energy, the Office for Management and Budget and the US State Department. Overall we have helped design more than 650 discrete innovation contests on the TopCoder platform for NASA and its partners.

The laboratory's work has shown that innovation contests can be routinely used to solve computational problems within the federal government and at elite academic medical centers. These problems can range from the design and development of robust software systems to the resolution of complex computational algorithm problems faced by engineers and scientists.

The development of a complex multi-state, multi program information technology solution for CMS serves as an example for the cost, speed and quality results arising from the use of innovation contests. The CMS program served to create a new software application suite that would assist in screening and registering health care providers for state run Medicaid programs. The aim was to better facilitate the screening of health care providers while at the same time lowering the burden on providers and reducing administrative and infrastructure expenses for states and federal programs. Ideally, this application would be able to ease provider enrollment processes while also identifying and preventing "bad actors" from enrolling as providers in state Medicaid programs and thus reduce fraud. The system also had to be backwards compatible with existing legacy systems and use modern shared and cloud-based information technologies. The CIL, with TopCoder, ran more than 140 contests, involving 1500 participants from over 35 countries to develop the application within nine months. Quality of the solution developed was judged to be above the standards typically followed by the traditional IT contractors, and cost analysis by CMS program managers revealed that a comparable system from a traditional vendor would cost $6 million as compared to the $1.5 million in charges to develop through innovation contests. Furthermore, the administrative cost of running and supporting a traditional procurement system was estimated at $1.4 million, as compared to less than $90,000 for the contest model. Thus the overall difference in cost, as estimated by CMS staff, was estimated to be on the order of 4.9 million (Garner and Wood 2013).

Similar cost, speed and performance gains have been achieved with CIL's work on the development of solutions to complex computational algorithms. The laboratory has completed 15 challenges in life sciences, space sciences and advanced analytics. Thirteen of the 15 challenges achieved their objectives by developing solutions that either met or vastly exceeded the comparative gold standard technical performance benchmarks that existed in the field. Two challenges failed to create satisfactory solutions. Table 1 provides an overview of the algorithmic challenges completed. The challenges typically delivered working solutions within several weeks and typically cost between $25,000 to $100,000 including reward money, platform fees and internal staff time. Equivalent effort within the host organizations would typically involve at least one post doctoral fellow and a principal investigator working on the problem for several months if not more.

The volume of innovation contests conducted through the laboratory allows for the occasional development and execution of an innovation field experiment. The CIL has exclusively focused on using algorithmic challenges as the vehicle for the field experiment. These experiments involve close collaboration, coordination and interaction with relevant scientific staff to ensure that a suitable problem statement can be developed so that contestants will be able to develop solutions. Simultaneous to the technical development is the establishment of the social science objectives and the experimental design. The experimental design drives the changes that we need to make to the TopCoder platform to ensure that the scientific objectives are met. Some of the changes we have put through the platforms included isolating treatment rooms to reduce threats to randomization, isolating communication amongst and to members, varying incentives, establishing team structures and team coordination, implementing various survey instruments and enabling search and matching amongst members.

The relationship with HMS and Harvard Catalyst (HC) has also enabled the development of a second type of experiment that has been focused on answering fundamental questions around the generation and evaluation of scientific research grant proposals within the HMS context. Harvard Catalyst is the university-wide translational science center with a mission to drive therapies from the lab to patients' bedsides faster and to do so by working across the many silos of HMS. A large portion of the Harvard Catalyst budget and outreach efforts for translational medicine is to offer grant funding to scientists. These internal grant competitions provide an ideal setting to investigate core research questions in the economics of innovation because researchers have to compete to win grants, these competitions involve evaluations and team collaborations dominate. The work with Harvard Catalyst has involved "layering" on social science randomization within the context of their grant-making. This has involved workshops with the relevant staff to help them understand the social science research objectives and ways in which Catalyst objectives are to be met. During an experiment, CIL staff work hand-in-hand with Catalyst staff and scientists to manage the entire grants process. Table 2 provides an overview of eight large-scale field experiments designed and executed by the laboratory over the past five years.

III Experimental Framework and Findings

The CIL has focused its efforts on understanding how individuals engaged in innovation-related problem solving efforts respond to exogenous treatments. Innovation contests form the basic experimental unit through which the laboratory's research investigations are organized. The laboratory designs and executes innovation contests in close collaboration with its partners to solve a particular innovation problem while simultaneously aiming to derive causal social science inferences. Partners seek assistance of the laboratory to have an innovation problem solved through an online contest (as in the case of NASA and the computational biology problems) or to run their own internal contest process (e.g. academic grant processes at HMS). In the first case, the steady stream of partner problems enable the laboratory to manipulate the online field setting of our online partner, TopCoder, to address specific research questions of interest, e.g. how might exogenously changing incentive types offered in a contest change participant behavior while a computational problem is being solved. In the second case, the particular objectives of our partner in running a grant competition inform the design of our research question, e.g. how collaboration might be encouraged in the generation of grant proposals for a particular research funding opportunity leads to an experiment in how exogenously shifting search costs for researchers impacts collaboration formation.

The economics of innovation literature provides the theoretical and empirical guideposts for the field experimental work undertaken by the CIL. In particular, the design of experiments undertaken by the laboratory has sought to understand how *incentives* and *knowledge* serve as inputs into the innovation production function and how the *search process* implemented during problem solving activity impacts various outcomes of interest:

1. **Incentives:** The basic building block within the innovation literature is that the provision of incentives influences the effort choices made by individual problem solvers. The incentive mechanism in a contest is the offering of a monetary prize in return for top performance. The prize motivates individuals to participate, and competition amongst participants to the claim the reward drives up effort performance. The presence of other individuals, however, complicates the effects of incentives on the actions taken by the individuals. The treatments investigated have included changing the level and types of incentives offered (pecuniary, job market signals, peer signals), modifying the rules of establishing winners and varying competitive intensity.

2. **Knowledge:** A perspective common amongst economists and management scholars is that innovation occurs when problem solvers develop solutions through novel (re)combinations of existing and new knowledge. Participants have access to an idiosyncratic stock of innovation-related problem solving knowledge, which is then used to develop new innovations and also to assess the value of these innovations. Experimental treatments have examined how varying

knowledge disclosure policies impacts participation in contests and how intellectual distance between expert evaluators and innovation proposals affects evaluation scores.

3. **Search Process:** The literature on innovation has productively cast the problem solving activity of innovators as a search for solutions under varying degrees of uncertainty. Innovators faced with an innovation problem, depending on the institutional context, can choose to problem solve on their own or with partners and also choose the direction of their search. The field experimental setup has enabled manipulating the institutional context to examine how preferences for working autonomously or with teams affect outcomes and how search costs shape the probability, type and productivity of team collaborations. In addition, the direction and shape of the search process itself has been studied in light of varying knowledge disclosure regimes.

Next we summarize four studies that highlight field experimental designs. The first two studies involved modifications to the TopCoder platforms and the second two involved redesigning the HMS grant process to achieve randomization and exogenous treatment deployment. We describe the substantive innovation problem, randomization approach and the results achieved.

Study 1: Knowledge Disclosures in Innovation

While there is plenty of anecdotal evidence that innovation contests can, under varying circumstances, sometimes outperform traditional (internal) modes of organizing innovation, direct comparative evidence is difficult to develop as researchers need to be able to examine performance simultaneously under both conditions. However, developing a direct comparison is important for both scholars, as we need to understand issues of economic efficiency and social welfare, and practitioners, who need to decide if innovation effort should be exerted internally or through external innovation contests. The need for research on this comparative question became apparent when HMS researchers approached us with the possibility of collaborating on understanding how external innovation contests could be used within the academic medical setting. As discussed in the introduction, we worked closely with HMS staff to identify a prototypical computational genomics problem that was challenging within life sciences and that could serve as means to compare contest versus internal performance.

Our paper (Lakhani et al. 2013), in collaboration with HMS researchers and TopCoder employees and participants, shows how over the course of two weeks, more than 122 solvers (out of the 722 individuals who initially signed up) from 89 countries created more than 650 solutions to the problem for a total prize purse of only $6000. The paper shows that thirty solutions exceeded by far the NIH and internal Harvard benchmarks and the best of them advanced the state of the art a factor of 1000. Figure 1 graphically illustrates the performance improvements. This was achieved via the contestants implementing 89 novel computational approaches to solve the problem, as compared to six approaches identified in the literature. In examining the solutions, our colleagues at HMS were simply taken aback by the sheer diversity of novel approaches used to attack the problem and the fact that none of the solvers had any background in computational biology.

The collaboration with HMS and TopCoder also provided us as opportunity to layer on a field experiment that investigated how knowledge disclosure policies may impact the rate and direction of innovative activity. One of the most basic distinctions between a contest and a community is the timing and form of knowledge disclosures during the problem solving process. Knowledge disclosures in a contest occur at the end when details of the winning solutions are made public. During the contest, participants typically work in secret and are not aware of their competitors' designs, and thus there is only *final disclosure*. In communities, the problem solving process is such that there is a continual sharing of knowledge about various solution approaches resulting in *intermediate disclosure*. This fundamental difference has direct effects on the rate and direction of inventive activity. Contests typically will create high incentives for individuals to participate and exert effort as they can appropriate all of the benefits of high performance for themselves. Meanwhile, intermediate disclosure in communities implies that participants will not have full appropriability, as others can use their discoveries for their own benefit, leading to depressed incentives.

At the same time, knowledge disclosures also impact the search process during problem solving. In communities, having access to the solutions, approaches and even mistakes of others can provide a significant boost to one's own problem solving effectiveness and can lead to convergence on the best approaches. Meanwhile in contests, one can expect search amongst contestants to be uncorrelated and potentially drive diversity in solution approaches. Testing the effects of varying knowledge disclosure policies on innovation poses significant research design challenges as causal inference requires that the problem to be solved, the profile and skills of the participants and the incentive schemes offered be held constant along with precise measures of innovative effort, performance and technological solution approaches developed.

Data analysis from the field experiment on the activities of 722 participants revealed that there were major differences in the effort, performance and search process implemented in contests and communities. The intermediate disclosure treatment directly led to lowered incentives in the form of fewer individuals choosing to get activated, and those that did participate exerted less effort as compared to under the final disclosure treatment. However, despite depressed incentives and participation, the intermediate treatment had higher innovative performance overall and on average. This can be explained by closely examining the solution approaches used by participants. Intermediate disclosure had the advantage of efficiently steering development towards improving existing solution approaches, which were already highly performing, limiting experimentation and narrowing technological search. Hence the disclosure policy can create altogether different effects on both incentives and search. We also found that the nature of the problem may be an important feature within the innovation contexts, as problems that may have a singularly maximal performance peak benefit more from intermediate disclosure approaches, while others that may have a rugged performance landscape will benefit from uncorrelated search. Thus disclosure policy is a fundamental organizing principle between contests and communities and more generally serves to inform the design of many of society's innovation approaches.

The results of this experiment prompted the TopCoder platform to offer a new type of contest structure that relied on participants initially working independently and then enabled them to use and borrow code from each other. At the same time, it allowed us to design a study that could obtain causal inference while comparing disclosure regimes that are typically occurring in very different empirical settings (e.g.: open source versus open science).

Study 2: The Impact of Sorting on Online Innovation Platforms

A crucial distinguishing feature of innovation contest models, as compared to the internal innovation process followed by most firms, is that it requires self-selection to create a match between the individual problem solver and the innovation challenge. While managers in a firm determine the tasks, incentives, and the organizational structure for their innovation workers, in crowd-based innovation, participants get to decide which tasks they are going to work on, the level of effort they are going to exert, which incentives will be most appealing, and if they prefer to work on their own or with others.

We investigated the importance of sorting by conducting a novel field experiment on how the ability to select one's preferred institutional regime for problem solving affects effort and performance in creating a solution to an innovation problem. Core to this study is the notion that sorting in the economy enables efficient allocation of talent and resources to important problems. A nascent literature in the economics of innovation and science has started to note that creative workers have certain institutional preferences, which drives their choices and effort (Stern, Cohen & German). This finding is also broadly consistent with theorizing by labor economists that differential incentive schemes sort and select worker effort and performance, primarily on the basis of skills.

We worked with NASA's Space Life Sciences Directorate to source an algorithmic problem from the space program and implemented its resolution as a contest on the TopCoder platform with a $25,000 prize purse and measures of effort and quality. Over 1000 software developers participated in our experiment over a 10-day period. Subjects developed algorithms to optimize the Space Flight Medical Kit for NASA's Integrated

Medical Model (IMM) software package. The problem specifically required participants to recommend the components of the space medical kit included in each space mission. The solution had to take into account that mass and volume are restricted in space vehicles and that the resources in the kit need to be sufficient to accommodate both expected and unexpected medical emergencies. The problem thus required a software solution that traded off mass and volume against sufficient resources to minimize the likelihood of medical evacuation.

We used this problem to design an experiment that enabled us to independently assess the impact of self-selection into a preferred work regime (i.e. working in a team or working autonomously) while *controlling* for skills and incentives. Our experiment was novel in the sense that instead of randomly assigning individuals to teams or solo competition treatments, we sought to elicit preferences from a subset of our subjects as to their choice of work regime. We implemented this selection experiment by rank ordering all subjects based on their prior TopCoder skill rating and then creating match pairs of individuals based on their skill. We then randomly solicited the institutional preference from one person in the matched pair and then assigned that same choice to the other person in the pair. Hence we had skill-controlled treatment and control conditions. We also randomized incentives in a way that some individuals were competing for cash prizes while others for fame and recognition. The solutions developed exceeded the benchmark NASA-developed solutions by both decreasing by an order of magnitude the time required to arrive at the recommendation and improving the potential simulated outcomes.

Our analysis (Boudreau and Lakhani 2011) found that allocating individuals to their preferred regimes had a significant impact on choice of effort level. Participants that chose the autonomous competitive regime, worked, on average, 14.92 hours compared to 6.60 hours, on average, for the unsorted participants. The effect was also positive and significant in the team regime, in which the sorted group worked, on average, 11.57 hours compared to 8.97 hours, on average, for the unsorted participants. Analysis of effort in terms of observable measures of code submissions revealed similar magnitude and significance as the hours of effort measure. We were also able to calibrate our results by showing that the effect size of the sorting mechanism was similar to the provision of pecuniary incentives in the autonomous competitive regime and about one third the value in the team regime. This experiment provides causal evidence for how an innovation worker's preferences for their work regime drives their effort choices and shows that the selection and sorting effects of our institutions for innovation (e.g.: garage startups, academic science, large firms, open source, innovation contests) are as salient as their treatment effects.

Study 3: Intellectual Distance and the Evaluation of Scientific Ideas

Essential to the innovation process is the selection of ideas that should be given resources and further developed while halting work on less promising proposals. Society expends considerable efforts towards the evaluation task. Inside organizations, executives have to choose between multitudes of competing proposals (some field reports note that over 3000 ideas are examined before a market entry is selected) and national funding bodies in the US allocate their billions of annual funding to an expert peer review process that involves thousands of scientists. A similar evaluation challenge exists for innovation contests that do not have access to a computer-based scoring and evaluation function; contest sponsors then have to rely on experts to help select amongst the contest entries.

A project with HMS on generating research proposals for evaluating the outcomes of a Type-1 diabetes research hypothesis-generation grant process provided the occasion to design a field experiment to understand how a relatively large panel of experts evaluate proposals that are close and/or distant to their own knowledge bases. A prior project with HMS and InnoCentive (a science problem contest platform) had generated 150 proposals that needed evaluation. Given the diversity of topics within the proposals (e.g., causes, prophylaxis, biological mechanisms, treatments and care), it became apparent that a broader range of scientific experts would be needed to helps select the best proposals. This issue became a research opportunity for the CIL to design a field experiment that could potentially answer important questions about how experts evaluate scientific ideas. Extant literature in the natural sciences has mostly raised issues of *ad hominem*, structural, social and political factors as driving scientific committee evaluations. We were interested in understanding how the intellectual distance between an expert evaluator and proposals affected scores, while controlling for quality and other factors. We were able to recruit 142 faculty members from Harvard Medical School to help us evaluate the proposals. Each evaluator assessed 15 randomly assigned proposals and each proposal received approximately 15 scores from randomly allocated evaluators, generating 2130 proposal-pairs. The proposal process was "triple blinded" in a sense that the identities of submitters and evaluators were blinded to each other, and that evaluators were not aware of each other.

Our analysis shows that knowledge-based biases significantly affect evaluation outcomes (Boudreau et al *forthcoming*). Access to fine-grained data on submitters and evaluators from HMS, in combination with analysis of the entire medical literature (via PubMed), allowed us to construct measures of *evaluator distance* for each proposal (the degree of overlap between an evaluator's knowledge (through their publications) and the knowledge in the proposals) and *proposal novelty* (the degree to which a proposal recombined knowledge in ways that were not present in the entire previous literature). We found that the closer an expert was to the field of the proposal, the harsher (more negative) their evaluation. We also found that the more novel a proposal, i.e. the more it contained novel recombination of existing knowledge that had not been published previously, the worse scores it received. Figure 2 shows the relationship between evaluation score and expertise distance and novelty graphically. The magnitude of these effects is such that they easily knock proposals from contending for funding. Our analysis of the data led us to ascribe these results to limits to human cognition, implying a bounded rationality explanation for the effects. Our paper is able to rule out as explanations both concerns about private (strategic) interests of evaluators and intellectual distance simply generating more noise in evaluations.

The findings of our paper have broad implications for how resources in the sciences are allocated (over $40 billion is annually allocated by the NIH and NSF) and provide explanations for concerns in the scientific community about incrementalism. Furthermore our paper shows how contest evaluation processes should be designed and potentially rectifies biases that may occur through various types of voting mechanisms.

Study 4: Search Costs in Collaborator Matching

Collaboration has become the default approach to knowledge production in the academic sciences (Jones XXXX). The formation of collaboration amongst scientists can be viewed as a search and matching process involving individuals who are seeking to find partners for a joint project. This feature can be enabling and constraining in academic science as researchers may face significant frictions in finding their most optimal collaborators. Harvard Catalyst worked with the CIL to help create a community of clinicians and scientists in the area of advanced medical imaging and to elicit innovative new research proposals in this area ($800,000 grant budget). Catalyst executives in particular wanted to ensure that the geographically and intellectually disparate researchers within HMS would be able to find each other and potentially collaborate on research projects. This created the opportunity to layer a field experiment on top of the grant competition to examine if search costs were economically significant in shaping the rate and type of scientific collaborations.

Our experimental manipulation involved exogenously shifting the search costs for random pairs of researchers and observing the impact on collaborations through application data on grant proposals. To run the experiment, we first needed to determine the potential "at-risk" population of researchers that might be interested in this grant proposal. We accomplished this by imposing a requirement that all interested researchers had to submit a statement of interest, outlining their research questions in advanced imaging, prior to the grant proposal deadline. This generated more than 400 applications from individual scientists across Harvard University and affiliated hospitals. We then required the applicants to participate in a face-to-face advanced imaging symposium at Harvard's Innovation Lab to learn more about the scientific advances in medical imaging, meet other researchers and understand the details of the grant application process. This was a first ever event to build a research community across scientific disciplines to address a common research problem at Harvard University. The advanced imaging symposium became the site for our randomized intervention on shifting search costs for researchers. The symposium was held over three consecutive evenings and applicants were randomly assigned to attend each evening. The symposium program was designed to have a common scientific and administrative introduction followed by information sharing sessions in separated "mixing" rooms.

We created standardized information sharing posters for each of the 400 applicants, based on their statement of intent, and randomly assigned them to one of four mixing rooms. Participants in each of the rooms got to meet and discuss with each other (between 20-40 researchers per room) their research interests and their specific research projects for 90 minutes. Thus we randomly reduced information search costs for partners for the scientists in each of the rooms while keeping them higher for all other scientists in the other rooms in the same night. Applications were due a month after the symposium and my laboratory also helped in creating a large-scale grant evaluation process. The grant process itself yielded more than 180 applications, far above the expectations of the Harvard Catalyst executives.

We found significant search costs with respect to finding collaborators (Boudreau et al. 2014). The treatment (a 90-minute structured information sharing session) increased the baseline probability of grant co-application of a given pair of researchers by 75% (increasing the likelihood of a pair collaborating from 0.16 percent to 0.28 percent). The magnitude of this effect is quite large, approximately one-third the effect of being from the same hospital (geography and institution) and in the same clinical area (intellectual domain). Further analysis showed that effects were higher among those in the same clinical specialization. The findings indicate that searching and matching between scientists is subject to considerable frictions, even in the case of geographically proximate scientists working in the same institutional context with ample access to common information and funding opportunities. The success of the advanced imaging symposium has prompted Harvard Catalyst to investigate and implement other such mixer activities for scientists at Harvard University.

Taken together, these studies combine to show how a systematic program for field experimentation can solve specific technical and organizational issues related to innovation while simultaneously allowing for the investigation of the underlying causal mechanisms that impact relevant outcomes. Next we discuss how the field setting influences the key design parameters available to a researcher considering embarking on innovation field experiments.

IV Innovation Field Experiment Design Parameters

The experience of the laboratory with running multiple innovation field experiments has yielded insights on the design parameters that require significant attention by researchers. These design parameters include selecting the appropriate innovation tasks to study and being attuned to the outcome measures, offering the appropriate stakes for the tasks to be accomplished and ensuring that randomization procedures are adhered to and executed by the researchers.

Innovation Tasks

Core to the innovation process is problem-solving activity by innovators to overcome some technological challenge. Field experiments in innovation thus need their subjects to engage in meaningful and relevant tasks that include generating solutions to problems and the evaluation of innovation proposals and projects. The core innovation tasks thus become the source of outcome measures to be used within the research programs. Problem solving related outcomes parameters include assessing technical performance, effort and collaboration-formation as the salient outcomes. Subjects in the experiments are all asked to solve a particular innovation problem and the technical performance of their problem solving effort is then assessed. Performance for computational problems is derived through an automated system that numerically scores competition entries and is developed through close cooperation with partners. If automated scoring is not available, then subjective expert evaluations form the basis for performance and are frequently used for grant proposal assessments. Effort exerted in creating a solution is also used as a relevant outcome. This can be based on observational data, as in a count of the number of solution attempts made during problem solving or the problem solving activity recorded on an online platform. Effort can also be obtained through survey data where the subjects are asked to report the hours invested in creating solutions. Depending on the research question, another outcome variable of interest is the formation of collaboration to solve a problem. In particular, the laboratory has framed collaboration for innovation as a matching problem where individual problem solvers are searching for a partner. Thus, achieving a match and engaging in collaborative effort also becomes a useful outcome variable.

Organizations expend significant managerial and expert review resources to assess the veracity of competing innovation proposals. However, how funders, expert reviewers, executives and sponsors evaluate innovative ideas has not been rigorously studied within the innovation literature. In many instances objective measures for the quality assessment of an innovation proposal is not available *ex ante* thus resulting in subjective assessments of quality. Study 3, discussed above, illustrates how evaluation within the innovation process can be amenable to field experimental approaches. Given the ubiquity of subjective evaluations there appear to be significant opportunities to apply the experimental took kit to this important aspect of the innovation process.

Stakes

The field experimental literature has identified the importance of relevant stakes for participating subjects (List 2009). Participants must be rewarded equivalently in the experiment as they may experience in natural settings without experimental treatment. In the case of innovation, this requires access to significant resources to ensure that the stakes offered match the general expectations that individuals have towards accomplishing the various innovation tasks that are on offer. Thus as Tables 1 and 2 indicate, partnership with host organizations that have real innovation problems needing to be solved along with a willingness to contribute funds for their resolution is necessary to ensure that appropriate stakes are being offered. As the tables illustrate, projects organized by the lab range from several thousand to tens of thousands of dollars for online computation challenges (roughly mirroring the market rates set by the TopCoder platform), to ensuring that hundreds of thousands of dollars are available in grant funding for prospective researchers. Stakes are thus important credibility signals to the subject pool that the researchers are serious about the outcomes and are conducting studies that will eventually result in some type of innovative output.

Randomization

Perhaps the most challenging and crucial aspects of conducting a field experiment is ensuring randomization of treatments to the recruited subjects and relative isolation between treatment and control subjects. While intuitive to researchers, the logic of randomization can cause significant consternation amongst sponsors and executives ultimately responsible for innovative outcomes. The first priority that the sponsors note is that the outcomes of the innovation task not be compromised due to experimental conditions. The natural inclination for many administrators is to "stack the deck" in a way to ensure success. This can subtly imply that the most highly skilled individuals for are assigned the treatment. It is thus critical for researchers to explain in layman's terms why randomization is critical to the success of an experiment. Researchers also need to explain the rationale for isolating treatments and preventing any spillover of information about the various treatment arms being run. Finally researchers need to ensure that the randomization procedure is completely within their control and that they have ensured that any threats to randomization are minimized. This aspect is potentially the most time consuming part of the experimental design, as sponsors need to be continually and repeatedly reassured that randomization is at the heart of the experiment and without it there is no possibility of causal inference.

V. Conclusion

The literature on contests is probably one of the most well-advanced and sophisticated theoretical subfields within economics (Konrad 2009). Over the last decade contest theory has escaped journal articles and textbooks and has been implicitly implanted within several large-scale innovation platforms that routinely offer contests as the primary incentive scheme to hundreds of thousands of participants. In addition, academic funding mechanisms can also viewed within the contest framework. This provides a unique opportunity for economics of innovation scholars to deploy field experimental methods to answer questions on both the optimal design of innovation contests and the general workings of innovation systems. Field experiments have the potential to provide unambiguous causal evidence on innovation topics while simultaneously assisting organizations with their innovation problems. We encourage our colleagues to complement their existing econometric-driven empirical research with exploring how to deploy field experiments on questions of their own interest.

References

Bandiera, Oriana, Iwan Barankay, and Imran Rasul. "Field experiments with firms." *The Journal of Economic Perspectives* (2011): 63-82.

Bannerjee, Abhijit V., and Esther Duflo. "The Experimental Approach to Developmental Economics." *Annual Reviews of Economics*. (2009): 151-178.

Boudreau, Kevin J., Nicola Lacetera, and Karim R. Lakhani. "Incentives and Problem Uncertainty in Innovation Contests: An Empirical Analysis." *Management Science* 57.5 (2011): 843-863.

Boudreau, Kevin J., and Karim R. Lakhani. "How to manage outside innovation." *MIT Sloan Management Review* 50.4 (2009): 69-75.

Boudreau, Kevin J., and Karim R. Lakhani. "Using the crowd as an innovation partner." *Harvard Business Review* 91.4 (2013): 60-69.

Boudreau, Kevin J., and Karim R. Lakhani. ""Open" disclosure of innovations, incentives and follow-on reuse: Theory on processes of cumulative innovation and a field experiment in computational biology." *Research Policy* 44.1 (2015): 4-19.

Brunt, Liam, Josh Lerner, and Tom Nicholas. "Inducement prizes and innovation." *The Journal of Industrial Economics* 60.4 (2012): 657-696.

Harrison, Glenn W., and John A. List. "Field experiments." *Journal of Economic Literature* (2004): 1009-1055.

Jeppesen, Lars Bo, and Karim R. Lakhani. "Marginality and problem-solving effectiveness in broadcast search." *Organization Science* 21.5 (2010): 1016-1033.

Jones, Benjamin F. "The burden of knowledge and the "death of the renaissance man": Is innovation getting harder?." *The Review of Economic Studies* 76.1 (2009): 283-317.

King, Andrew, and Karim R. Lakhani. "Using open innovation to identify the best ideas." *MIT Sloan Management Review* 55.1 (2013): 41-48.

Lakhani, Karim R., Kevin J. Boudreau, Po-Ru Loh, Lars Backstrom, Carliss Baldwin, Eric Lonstein, Mike Lydon, Alan MacCormack, Ramy A. Arnaout, and Eva C. Guinan. "Prize-based contests can provide solutions to computational biology problems." *Nature biotechnology* 31, no. 2 (2013): 108-111.

Lerner, Josh, and Jean Tirole. "Some simple economics of open source." *The journal of industrial economics* 50.2 (2002): 197-234.

List, John A. "An introduction to field experiments in economics." *Journal of Economic Behavior & Organization* 70.3 (2009): 439-442.

List, John A. "Why economists should conduct field experiments and 14 tips for pulling one off." *The Journal of Economic Perspectives* (2011): 3-15.

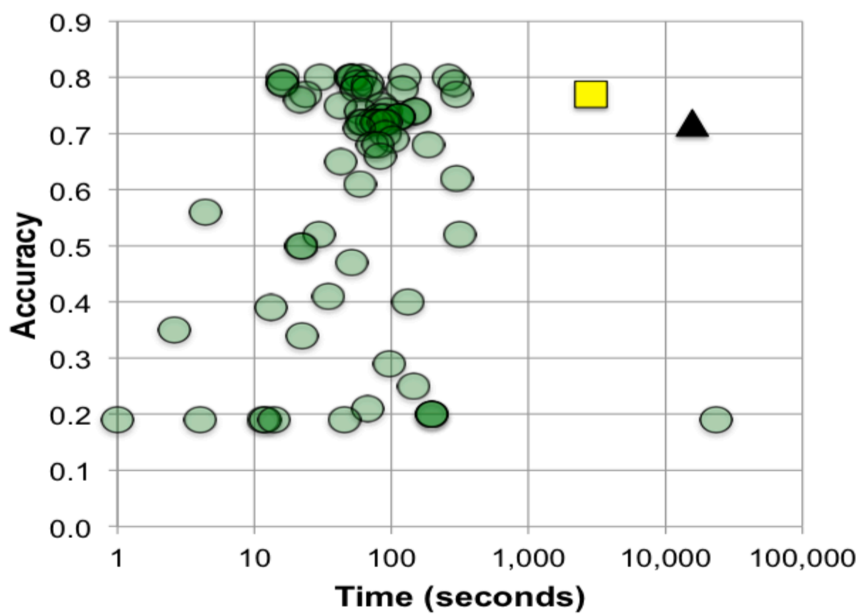
Murray, Fiona, Scott Stern, Georgina Campbell, and Alan MacCormack. "Grand Innovation Prizes: A theoretical, normative, and empirical evaluation." *Research Policy* 41, no. 10 (2012): 1779-1792.

Samuelson, Paul, and William Nordhaus. *Principles of economics.* New York: McMillan (1985).

Scotchmer, Suzanne. *Innovation and incentives*. MIT Press, 2004.

Williams, Heidi. "Innovation Inducement Prizes: Connecting Research to Policy." *Journal of Policy Analysis and Management* 31.3 (2012): 752-776.

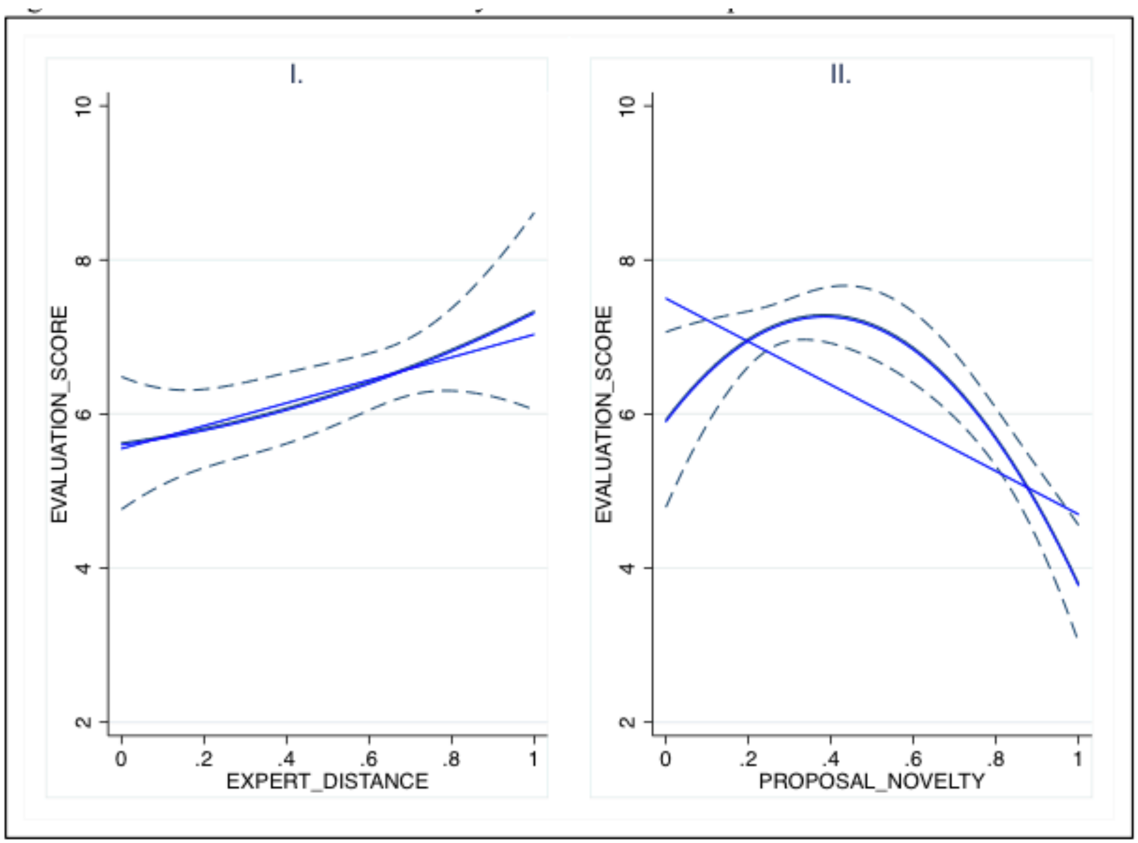
Figure 1 Accuracy score vs. speed of contest-commissioned immunoglobulin sequence annotation code



Source: Lakhani et al. 2013

Note: Circle represents contest entry. Square is Harvard code. Triangle is NIH MegaBlast code.

Figure 2 – Impact of intellectual distance on evaluation scores



Note. 90% confidence intervals shown.

Table 1: Computational Algorithmic Challenges Completed by Crowd Innovation Laboratory

Challenge	Sponsor	Prize Amount	Number of Contestants (Submissions)	Time Length (Days)	Performance Results
Computational Biology					
Antibody Sequencing	HMS	$6,000	122 (654)	14	Exceeded benchmark results from HMS and NIH. Three orders of magnitude improvement.
Classification of Minority Variants in Pooled HIV Sequencing	HMS	$5,000	196 (668)	14	Exceeded HMS benchmarks. Classification now possible at 0.1% versus previously at >0.5%.
Antibody Clustering	Scripps /NASA	$8,500	40 (214)	7	Exceeded Scripps benchmarks. Four orders of magnitude improvement.
Knowledge Extraction via Natural Language Processing for PubMed Articles	Scripps/NIH /NASA	$30,000	82 (1700)	8	Exceeded Scripps/NIH "F-measure" (precision and recall) improved by 5%.
Chemical Toxicity Prediction	EPA /NASA	$10,000	47 (783)	14	Improve EPA internal model by 20%.
Cyano Bacterial Modeling	EPA/NASA	$15,000	30 (460)	21	Source problem, develop scoring algorithm, modify TopCoder platform & deliver solutions.
Aerospace Sciences					
International Space Station Longeron	NASA	$30,000	459 (2009)	21	Design and develop innovation competition for MGH.
Asteroid Data Hunter I	NASA	$10,000	60 (301)	14	Reduce false positives by an order of magnitude.
Asteroid Data Hunter 2	NASA	$20,000	47 (256)	14	Increased asteroid detection by 15% compared to benchmark algorithm.
Asteroid Tracker	NASA	$15,000	43 (299)	14	Met current benchmarks and established proof of concept for algorithmic performance
Planetary Data Systems – Saturn Cassini Mission	NASA	$25,000	15 (255)	14	New algorithm to detect propeller objects in the rings of Saturn. Identification of objects up by 30% with 80% accuracy.
Satellite Image Detection	NASA/UCSD	$15,000	39 (357)	21	Reduced need for human labeled data and matched manual performance.
Advanced Analytics					
Image and Text Analysis in Patent Documents	NASA/US PTO	$50,000	140 (1797)	30	*De novo* algorithm for automated detection of patent images, parts and related text.
Predicting Probability of Atrocity Events using News Data	NASA / USAID	$25,000	93 (592)	21	*De novo* algorithm that outperforms naive frequency based predictions of human rights violations by 60%.

Table 2: Innovation Field Experiments by Crowd Innovation Lab at Harvard Institute for Quantitative Social Science

Sponsor / Platform	N	Innovation Objectives	Research Questions	Key Challenges
1.HMS / TopCoder	722	Develop sequence alignment algorithm for genomics application ($6,000 prize pool).	How do disclosure regimes impact the rate and direction of innovative activity (contests versus communities)?	Source problem, develop scoring algorithm, modify TopCoder platform & deliver solutions.
2.NASA / TopCoder	1200	Develop algorithm to create most optimal space medical kit for long-term space journeys ($25,000 prize pool).	How does self-selection into autonomous work versus team production drive effort and productivity?	Source problem, develop scoring algorithm, modify TopCoder platform & deliver solutions.
3.HMS / InnoCentive	294	Generate and evaluate new research for treating Type 1 Diabetes by engaging Harvard and rest of world ($30,000 & $1,000,000 in grant funding).	How can innovation contest mechanisms be applied to academic medical centers? How does evaluator expertise and knowledge impact the scoring of frontier science projects?	Design and execute a new grant process that enables new participants to contribute. Develop and execute a randomized and triple-blinded evaluation process that enables grants to be awarded.
4. HMS / HBS iLab	450	Encourage scientific proposals in advanced medical imaging across Harvard and help facilitate new collaborations ($800,000 in grant funding).	How do search costs impact the formation of scientific collaboration? How do peer reputation incentives impact scientific effort?	Design and execute an end-to-end new grant process that can build imaging community across Harvard. Identify & qualify population of potential participants, design & administer randomized information sharing sessions at HBS iLab, establish and coordinate grant submission requirements, drive evaluation of proposals.
5. NASA / TopCoder / Google	1000	Develop algorithms for autonomous space transportation robots ($35,000 prize pool).	What is the role of explicit peer and job market signals as compared to pecuniary incentives in a contest setting?	Source problem, develop scoring algorithm, modify TopCoder platform & deliver solutions. Attract Google and NASA JPL laboratory as sponsors to generate job market signals.
6. US Patent Office / TopCoder	1000	Develop image and text detection algorithms for US Patent Office ($50,000 pool).	What are the costs and benefits of self-organization as opposed to centralized assignment into teams?	Source problem, develop scoring algorithm, modify TopCoder platform & deliver solutions.
7. HMS / MGH	350	Create an internal contest for MGH Cardiac Center staff to generate innovation proposals.	How do extrinsic, intrinsic and pro-social incentives drive participation and effort in an internal solution generation contest?	Design and develop innovation competition for MGH.
8. NASA / Scripps	299	Improve NIH natural language processing algorithms.	How do races and tournaments differ?	Source problem, develop scoring algorithm, modify TopCoder platform & deliver solutions.