

Innovation Experiments: Researching Technical Advance, Knowledge Production and the Design of Supporting Institutions

Kevin J. Boudreau
Karim R. Lakhani

Working Paper 16-001



Innovation Experiments: Researching Technical Advance, Knowledge Production and the Design of Supporting Institutions

Kevin J. Boudreau
Harvard Business School

Karim R. Lakhani
Harvard Business School

Working Paper 16-001

Copyright © 2015 by Kevin J. Boudreau and Karim R. Lakhani

Working papers are in draft form. This working paper is distributed for purposes of comment and discussion only. It may not be reproduced without permission of the copyright holder. Copies of working papers are available from the author.

Innovation Experiments: Researching Technical Advance, Knowledge Production and the Design of Supporting Institutions*

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

July 2015

Forthcoming in Innovation Policy and the Economy, volume 16, edited by Josh Lerner and Scott Stern, NBER, 2015

Abstract: This paper discusses several challenges in designing field experiments to better understand how organizational and institutional design shapes innovation outcomes and the production of knowledge. We proceed to describe the field experimental research program carried out by our Crowd Innovation Laboratory at Harvard University to clarify how we have attempted to address these research design challenges. This program has simultaneously solved important practical innovation problems for partner organizations, like NASA and Harvard Medical School, while contributing research advances, particularly in relation to innovation contests and tournaments.

***Acknowledgements:** The experimental research described in this paper has been accomplished at the Crowd Innovation Lab | NASA Tournament Lab at Harvard University's Institute for Quantitative Social Science. We are grateful for support from NASA's Human Exploration and Operations Mission Directorate for their support in establishing the lab. In particular, Jason Crusan and Dr. Jeffrey Davis helped create the rationale for the integration of a research program on innovation with NASA's operational technology needs. Further support from NASA's personnel including Karl Becker, Lynn Buquo, Mike Ching, Carol Galica, Steve Rader, Allison Wolff, and Carolyn Woolverton, has been invaluable with lab operations. The following TopCoder executives and managers (current and former) provided platform access, experimental support and advise on community relations: Ambi Del Villar, Mike Epner, Jack Hughes, Robert Hughes, Ira Heffan, Tim Kirchner, Andy LaMora, Mike Lydon, Dave Messenger, Mike Morris, Jennifer Odess, Rashid Sial, and Narinder Singh, and Ashley Steele. Harvard Medical School's Dean for Clinical and Translational Research, Professor Lee Nadler, leadership of the Reactor Program, Harvard Catalyst, Professors Eva C. Guinan and Anthony Hollenberg and Program Director Gary Gray were essential partners in establishing field experimental projects in an academic research setting. Over the years we have been privileged to have the following members of our lab participate with us on our various projects: Michael Menietti, Andrea Blasco, Annika Mueller, Christoph Reidl, Ina Ganguli and Partick Gaule. The lab is managed by Jin Paik and Rinat Sergeev serves as senior data scientist. The HBS Division of Research and Faculty Development and the London Business School have also supported the work described in this paper.

1 Introduction

In economics and many other social sciences, the most common empirical research approach is one of studying natural data obtained from observational methodologies, analogous to those collected and studied by astronomers and meteorologists (Samuelson and Nordhaus 1985, List 2009). Nonetheless, even in using naturally occurring data, researchers are typically interested in estimating “treatment” effects and causal relationships in a population of interest. The data provided by nature might, in this regard, simply be understood as the result of an inadequately designed experiment, with econometric methods being a means of “correcting” for these imperfections, so as to restore an ability to estimate causal relationships and to identify relevant model parameters. This has been achieved with varying degrees of success in the profession (Heckman and Leamer 2007, Imbens and Wooldridge 2008). At least in this sense, the regular practice and goals of empirical economics researchers might be understood as closely allied with those of more formal experimental research (Harrison and List 2004).

However, whereas the use of formal randomized controlled trials has been the norm over the last 100 years in many areas of natural science, such as chemistry, physics, biology and medical research, prior paradigmatic views about the potential of economists to run more formalized controlled experiments were quite negative for many decades (Samuelson and Nordhaus 1985, List 2009). Only in the last two decades has there been an embrace of laboratory and field experimental approaches within the profession (List 2011). This rise was concurrent with and perhaps not unrelated to growing interest in behavioral economics, policy evaluation and rising standards and concern for attaining causal inferences more generally in the profession.

Experimental approaches have by now gained the most substantial ground in sub-disciplines including behavioral, developmental and labor economics (Bandiera et al. 2011, List 2009, 2011). By contrast, the economics of innovation literature has generally lagged behind in adopting an experimental approach. In this paper we argue that the nature of the innovation process and supporting institutions lead to five unique challenges that raise significant barriers for the design and execution of innovation experiments. These challenges include: 1) The nature of the knowledge production function, 2) The unit of analysis, replication and sample size, 3) Selection and/versus treatment effects, 4) Designing (organizational) treatments and counterfactuals and

5) Representativeness, validity and fine-grained measures. We then proceed to discuss our research program undertaken within the framework for the Crowd Innovation Laboratory 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 particular focus area of the Laboratory has been the design of innovation contests and tournaments, such that real-life technological problems are addressed while manipulating and measuring features of the innovation process. This leads us to a description of the approaches we have taken to attempt to address or mitigate innovation experiment challenges through the illustration of three examples of field experimental research designs.

2 Challenges of Designing Innovation Experiments

Whatever the field or domain, embracing experimental methods requires significant commitment and investment by the researcher (see, for example, Duflo and Banerjee 2009, Bandiera et al. 2011; List 2011). In addition to developing a theory-based perspective on an appropriate research question, scholars need to translate the theory to the experiment through a precise formulation within the observed measures and controls, design and implement exogenous treatments, obtain institutional review board approval, recruit representative subjects, ensure treatment and control isolation, record data, and then proceed with data analysis. Further, in the case of field experiments, researchers may require considerably more substantial resources in implementing and carrying out research designs. They need to gain cooperation from organizations in the field (e.g. an organization that provides access), assure that the research design is of interest to the organizational hosts, ensure that there will not be any harmful effects on the organizations' operations and communicate results and feedback to the sponsor. These and endless other fine points create considerable "entry barriers" and explain, in part, the slow adoption of experiments in economics and certain other social sciences.

Despite such barriers, as remarked above, considerable progress has been made in a range of fields—and particularly in labor economics, development economics, behavioral economics and consumer behavior. Experimental research in innovation, particularly research that studies how the design of relevant underlying institutions and organizations influences performance and

efficiency outcomes, lags considerably behind. Here we review five of the most conspicuous challenges that face empirical researchers working on these questions. These points will then form the basis for describing the particular approaches we have taken in a handful of research projects in the illustrative discussion following this section.

2.1 Multiple Mechanisms Shaping Innovation – and the “Knowledge Production Function”

If the theory required to explain any and all innovation phenomena and predicted patterns were restricted only to the structure of incentives, there might be less of a gap separating research on innovation and the production of knowledge versus that in other fields of economics. We might in such an instance simply define all desirable outcomes in relation to a neoclassical production function and some desired benefit function and assess individual decision making in relation to private costs and strategic interactions.

Of course, the basic building block within the economics of innovation literature is that the provision of incentives influences the effort choices made by individual problem solvers (e.g. Holmstrom 1989, Aghion and Tirole 1994, Scotchmer 2004, Lerner and Wulf 2007, Manso 2011). However, innovation only occurs when problem solvers develop solutions through novel (re)combinations of existing and new knowledge (Schumpeter 1934; Romer 1990; Weitzman 1998). Participants have access to an idiosyncratic stock of innovation-related problem solving knowledge, which is then used to develop new innovations as well as to assess the value of these innovations. This process is also deeply shaped by causal ambiguity in precisely how knowledge is produced (Rosenberg, 1982; Jones 2009). This ambiguity is associated with extraordinary variation in inherent productivity of different individuals and organizations attempting to solve innovation problems and is also quite closely related to the sometimes profound uncertainty regarding the appropriate paths and skills and approaches required to advance on a given problem (e.g., Boudreau et al. 2011). Within a context of cumulative and recombinant advances, where paths and relevant skills are not nearly always clear, there is something of a “search” process that takes place (Nelson and Winter 1982). In a context of multiple actors, recombinant cumulative search and high uncertainty, a number of other mechanisms are also often implicated in the process by which inputs are converted to outputs,

including sociological (e.g., Kuhn 1962, Azoulay 2004), psychological, (Amabile 1983) and cognitive (Boudreau et al. 2011). The stochastic nature of innovation outcomes can also often be a first-order aspect of the production of innovations, rather than simply a “residual error” (Nelson 1959).

Thus, in short, the “production function” underlying innovation and knowledge production is anything but straightforward, and attempts to study the innovation process and how the design of underlying institutions plays a substantial role in shaping the innovation process must confront the range of mechanisms playing a role. It is often difficult or not possible to artificially suppress the action of certain mechanisms when attempting to study the true nature of the process, capturing first-order determinants—even within an experimental set-up. This creates pronounced challenges in formulating designs to capture appropriate end points.

2.2 Unit of Analysis, Replication and Sample Size

The nature of the innovation process implies that technical advance and the production of new knowledge is often not just a problem of encouraging individual or group creativity, although that is certainly an important element of the wider picture of managing innovation. Given the unavoidable role of uncertainty, factors such as diversity and strategic interactions play a role within the processes of innovative recombination, cumulative advance and experimentation. Therefore, it is in many cases not sensible to attempt to artificially isolate a lone subject and consider impact of a varying treatment. For example, knowledge flows and accumulation are inherently problems involving multiple actors, as are the range of approaches of organizing innovation through some form of competition or collaboration.

Moreover, in many and perhaps most contexts, the choice to “enter” and engage in development in a particular area is itself hardly non-random and is often in direct response to expectations of the nature of institutions in which an innovation problem might be addressed, as when a scientist considers whether to work in academia or industry (Stern 2004), a technologist chooses whether or not to pursue a given problem or an individual solver chooses whether or not to join a given crowdsourcing platform to address certain problems.

Thus, the nature of the innovation process often implies that the appropriate unit of analysis is greater than the individual—frequently considerably larger: the population of innovators, or even the population of prospective innovators who choose whether to enter and pursue a problem when considering the nature of the institutions into which they will enter. The potentially large scale of relevant units of analysis can create notable challenges where large numbers of subjects may be required for just one observation, let alone multiple treatments and replication.

2.3 Selection versus/and Treatment Effects

Questions of treatment effects versus selection are longstanding issues—and often problems of econometric estimation—in all manner of empirical research in social science, as they are in research on innovation. The great contrast in experimental research—and its gold standard of randomized controlled trials—is that the *raison d’être* of these methods might even be said to erase the “problem” of selection effects, altogether. (That is not to say that estimating the treatment effect necessarily implies a clear interpretation of what might be complex and nuanced phenomena as in earlier in 2.1.)

The eradication or control of selection effects in experimental design, however, is not always entirely appropriate in studying innovation processes. The kinds of researchers, technologists and innovators who choose to work on one problem or another within a given institutional regime might itself be the single most important determinant of an institution’s character. We might also expect that the sort of individuals who are selected to a given environment (given responsiveness to incentives and the particular knowledge and ideas they embody, etc.) is highly salient in how any “treatment” effect might play out.

Of course our point here is not that selection and treatment effects need always be considered alongside one another and that randomized controlled trials no longer have a use. Rather, simply that the existing use of randomized controlled trials and its emphasis on treatment

effects can distract from the essential role of selection as a primary – and often the most important – determinant of innovative outcomes (cf. Lazear et al. 2012).

2.4 Designing “Institutional Design Treatments” and Counterfactuals

Work in experimental economics, particularly behavioral and labor economics, has made considerable progress through isolated psychological and situational manipulations to consider systematic behavioral responses. However, within the innovation context, where research is geared towards the design of institutions and organizations that create the conditions for problem solving effort exertion and the resultant technological and knowledge outcomes, the challenge to be confronted is that a complementary set of design variables need to be considered simultaneously (Milgrom and Roberts 1990).

It is hence rare that the most important dimensions of an organization or institution can be manipulated in isolation, as they are typically chosen with these other dimensions in mind. Take for instance that the patent system simultaneously mandates knowledge disclosures and monopoly rights over follow-on innovations. Thus any attempt to understand the impact of knowledge disclosure in innovation needs to consider both the benefits of positive spillovers and the incentive effects of monopoly provision.

To the extent that innovation research (or any other sort of research) relates to organizational and/ or institutional factors, the design of treatments must somehow address this challenge. Researchers can either artificially manipulate a single dimension within a treatment, so as to be scientifically parsimonious in the definition of a treatment, or else define a treatment in relation to variation of multiple design variables at once, so as to reflect distinct complementary combinations that might have greater salience to the research question and the innovation phenomenon.

2.5 Representativeness, Validity, and Fine-Grained Measures

On the one hand, innovation research might be quite narrow and test quite particular mechanisms that might not be specific to researchers or innovators. For example, where research involves testing narrow mechanisms related to say the socialization of small groups, or

say responses to particular sets of incentives, it might be possible to test results of theory with subjects in a synthetic lab environment. However, to the extent that research on innovation and supporting institutions is meant to discover and document what may be subtle, nuanced and yet not entirely understood mechanisms (as in 2.1) or to the extent that the responses of individuals depend quite particularly on the actors under study (as in point 2.3), the nature of problems they solve, the particular institutional details or combinations of institutional details (as in point 2.4), then the question of representativeness becomes all the more salient.

Therefore, a challenge in using experimental methods and apparatus is to take advantage of explicitly designed controlled environments. However, at the same time, there is a challenge of introducing sufficient representativeness so as to capture true mechanisms and to observe responses that calibrate to those that might be expected in the economy more broadly. This representativeness challenge requires close attention to the innovation tasks under study, the stakes to complete the tasks and the individuals engaging in problem solving activity.

Core to the innovation process is problem-solving activity by innovators to overcome some technological challenge. Experiments in innovation thus need their subjects to engage in meaningful and relevant tasks, which can 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. 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.

A related and an understudied aspect of the innovation process is the evaluation of innovation proposals. Organizations pushing the frontiers of knowledge expend significant managerial and expert review resources to assess the veracity of competing innovation proposals. In many instances objective measures for the quality assessment of an innovation proposal is simply not

available (or possible) *ex ante*, thus creating research challenges in understanding the drivers of evaluation.

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. 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 on offer and also attract the appropriate caliber of individuals. Stakes are 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.

3 Operationalizing Innovation Field Experiments via Industry Collaborations

The precursor to the establishment of the Crowd Innovation Laboratory (CIL) was our prior work in understanding the factors underlying performance in innovation contest platforms like InnoCentive (Jeppesen and Lakhani 2010) and TopCoder (Boudreau et al. 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 going beyond using observational data and conducting field experiments.

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 contests¹ 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

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

contests to induce 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's mission 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 to provide empirical estimates. The CIL's work has begun to rectify this gap and has taken canonical theories from the textbook to the field.

The stimulus to establish the laboratory came through 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.

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 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 also assisted other federal 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.

Table 1 provides an overview of the algorithmic challenges completed. The CIL has demonstrated significant gains in cost effectiveness, speed to solution and quality through the use of innovation contests. 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. The challenges typically delivered working solutions within several weeks and

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.

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 have 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 as 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.

4 Case Studies in the Design of Innovation Experiments

The earlier sections set forth a particular set of priorities and challenges in experimental research on innovation and the infrastructure and process needed to develop a robust innovation experiments platform. This might, in part, explain some of the lag of empirical research on innovation in embracing experimental methods. In this section, we present in the form of three case studies of our research projects, exemplifying possible approaches to overcoming these challenges, while making empirical progress. Table 3 summarizes the unique research design challenges, the associated solutions needed to accomplish the study objectives and the crucial role of close organizational partnerships in achieving our study objectives.

4.1 A Research Design to Investigate “Openness” and Knowledge Disclosures

This research project was designed to address two key empirical questions. First we wanted to understand how the outcomes of an innovation contest compared to the results of traditional grant and internal efforts. Second we wanted to investigate the role of varying knowledge disclosure policies in shaping incentives and the innovative search process. We designed an experiment to address both of these questions simultaneously.

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. We worked closely with HMS staff to identify a representative 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 demonstrates that thirty solutions exceeded by far the NIH and internal Harvard benchmarks and the best of them advanced the state of the art by a factor of 1000. Figure 1 (from Lakhani et al, 2015) graphically illustrates the performance improvements. This performance was achieved via the contestants implementing 89 novel computational approaches to solve the problem, as compared to six approaches identified in the literature. The resultant paper, published in a natural science journal (*Nature Biotechnology*) provided a general scientific audience with guidance towards understanding the underlying mechanisms about contest performance and addressing their concerns around representativeness through the selection of an appropriate research problem that was of general interest to the field.

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 (Boudreau and Lakhani 2015). Our own prior research on the use of crowds to solve innovation problems had identified contests and communities as two distinctive institutions that organize and shape participant efforts (Lakhani and Panetta 2007, Boudreau and Lakhani 2009, 2013). 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. In addition careful attention had to be placed on the appropriate unit of analysis, replication strategy and sample size. Since our main interest was in understanding the role of disclosure policies we chose to apply treatments to a large "population" of solvers that could then choose (or not) to act within the framework. This naturally limited replication options as we had to create large enough comparison groups that could allow us to infer the effects of the treatments on a range of prospective problem solvers – that would respond to the knowledge disclosure policies. Given the initial rate of signups to the problem on the TopCoder platform, 722, we created two main (equal sized) comparison groups under the intermediate and final disclosure regimes and a third "robustness testing" group that had the knowledge disclosure policies switch from final to intermediate during the experiment.

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 exert effort, and those that did

participate exerted less effort as compared to under the final disclosure treatment. (Note here, that conditional on the treatment, non-participation in the contest can be interpreted as an incentive effect of the 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).

4.2 A Research Design to Evaluate Non-Random Sorting onto 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.³

We investigated the importance of the sorting mechanism 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 drive their choices and effort (Stern 2004, Sauermann and Cohen 2010). 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 (Lazear 2000).⁴ Here we sought to understand how individual preferences for autonomous work or team production shaped the effort of participants.

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 and sorting into a preferred work regime (i.e. working in a team or

³ Of course workers initially sort into all types of organizations and jobs. However once inside the organizations they are under managerial control.

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 \$1000 cash prizes while others had no pecuniary incentives. The high number of individual signups allowed us to increase the replication of our treatments, with 6 to 7 groups of 20 individuals (or 4 teams of 5 individuals) working independently within the 2*2*2 design.

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. In addition 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.

4.3 A Research Design on the Impact of “Intellectual Distance” and “Novelty” in the Evaluation of Innovation Proposals

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 (Guinan et al. 2013). 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 help 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 (see Lee 2012, Lee et al. 2013). 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. Note that in most naturally occurring evaluation settings, evaluators typically select into assessing specific

proposals that they are either experts in or are assigned to committees that have a specific research topic in mind, hence limiting identification for questions around intellectual distance. The proposal process was also “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, a feature not commonly followed in grant processes but crucial to our research question.

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.

4.4 The Role of Organizational Partnerships

As the case studies make clear, embarking on innovation experiments requires developing a research infrastructure that can fundamentally take ownership for delivering innovation outcomes for sponsor organizations. In effect innovation experiments require “industrial scale” effort and resources for their successful design and execution in the following areas:

- **Sourcing problems:** The task to be solved needs to be relevant for innovation outcomes and calibrated to the requirements of the organization. Expertise in both identifying problems and creating appropriate scoring systems is needed as part of the core research team.
- **Modifying and (re)designing research environments:** Theory often places stringent demands on empirical data contexts to ensure that identification and causality can be inferred. Hence the organizational partners and contexts need to be flexible and fungible to accommodate research requirements. This includes making changes to platforms and internal processes to ensure that treatments can be executed and that the requirements of randomization, replication and participant isolation can be accomplished.
- **Access to a subject pool:** The sine qua non of innovation experiments is access to subjects that can partake in innovation-related problem solving. Organizational partners have to be able to provide the research team with a sufficient number of subjects that can undertake the innovation task at hand and provide detailed information about their skills, abilities and backgrounds to allow for experimental allocation to treatments and ex post controls during analysis and interpretation. The ability to host, interact and manage hundreds if not thousands of subjects in both online and offline settings is needed during the experiment design and execution phases.
- **Operational infrastructure, personnel and resources:** Innovation experiments also need to be actively managed from the design through the execution phase. The research team needs to be able to interface with sponsor organizations on an ongoing basis during the design phase to ensure that both the research objectives and the practical organizational

requirements are being met. The execution phase requires close monitoring of the experimental environment and flexibility and imagination in developing solutions to unforeseen circumstances. Researchers of course also need resources to fund the staff and temporary personnel needed to design and run the experiment and for “payments” for the innovation tasks accomplished by the subjects. As Table 2 shows, simply the “prize” funding for our experiments have ranged from \$6000 to \$800,000 and this has typically been achieved by the sponsor organization providing the funds.

Overall the design and execution of innovation experiments requires a very high degree of trust between the research team and the sponsoring organization. In our experience, achieving this trust entailed scoping out appropriate and achievable initial objectives that met the requirements of the sponsors, e.g.: initiating small pilot projects to prove that effective solutions can be developed and that the research teams have the ability to collaborate, and then following up with more ambitious projects. The initial acts of trust required investment on our part in time, effort and resources, without the guarantee that a viable research project could be developed.

5 Insights for Policy

Two sets of implications for policy implied by the earlier discussion include (1) how policymakers and researchers can fund, sponsor and use field experiments; and (2) the role of contests in innovation procurement. As we earlier argued and illustrated with examples, researching innovation comes with particular challenges. These create additional burdens for innovation research with experiments, although perhaps at the same time providing added returns. In particular, the research infrastructure and sponsor organizational relationships (and trust) required to execute innovation experiments is quite labor intensive and requires resources that are not commonly available through conventional funding arrangements. Although formal field experiment laboratories that tie practical outcomes to economic theory and analysis are increasingly common in development economics, to the best of our knowledge the CIL (and our related efforts in field experiments) is unique within the field of economics innovation.

However, given the importance of innovation and knowledge production to economic growth, it is incumbent on the scholarly, policy and funding communities to catalyze research that diligently develops a “science of innovation” that is empirically informed through field-based randomized control trials. Policy makers and government agencies can play an important role funding and supporting field experimental research on innovation and making more programmatic attempts to develop collaborations between empirical economists and government agencies pursuing innovation. Economists can also push towards creating research infrastructure that enables them to work hand-in-hand with organizations engaged in innovation activities to take a scientific approach towards applying various theoretical and practical levers that improve research productivity. Given the very large footprint of federal and industrial research and development activities there will be many opportunities for scientific collaborations between academia, government and industry on this topic.

Our research has also most practically shown that innovation contests can be used routinely to procure innovation and technology. The development of a complex multi-state, multi program information technology solution for Center for Medicare and Medicaid Services (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).

The policy challenge now is to understand how innovation contests can become a routine part of federal procurement for technology. The academic and policy establishments take for granted that traditional contractor-based procurement or research grants to academic institutions are the preferred modes for encouraging innovative outcomes. Our research approach has begun to provide some initial evidence that innovation contests may also be a viable option to sourcing technology. Although the America COMPETES Act provides a framework for US agencies to pursue prize-based contests in procurement, there is a significant lag between the spirit of the legislation and its adoption within the government agencies. Policy makers now have the opportunity to conduct systematic assessments of contests, grants and contracts as vehicles to drive innovation and can direct efforts so that the internal organizational resistance to a “new” way of procuring technology does not trump the importance of running careful counterfactual-based economic assessments.

6 Conclusion

Experiments have now become a standard approach to deriving insights in economics along with formal modeling and econometric analysis of observational data. In this paper we have sought to highlight the particular challenges faced by the economics of innovation literature to implement experiments. In particular we argued that the very nature of the innovation process and the surrounding organizations and institutions that support it raise non-trivial entry barriers to researchers interested in the experimental approach. We then provide the rationale and design of a systematic program and associated laboratory that has designed and executed innovation experiments within the context of contests in close partnership with NASA, Harvard Medical School, and the TopCoder online innovation platform. The laboratory’s mission is to solve our partners’ innovation problems while simultaneously undertaking randomized control trials on topics related to innovation contests.

The literature on contests is probably one of the most well-advanced and sophisticated theoretical subfields within economics. 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 be 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 exploration of how to deploy field experiments on questions of their own interest.

References

Aghion, Philippe, and Jean Tirole. "Formal and real authority in organizations." *Journal of political economy* (1997): 1-29.

Amabile, Teresa M. "The social psychology of creativity: A componential conceptualization." *Journal of personality and social psychology* 45.2 (1983): 357.

Azoulay, Pierre. "Capturing knowledge within and across firm boundaries: evidence from clinical development." *American Economic Review* (2004): 1591-1612.

Azoulay, Pierre, Joshua S. Graff Zivin, and Gustavo Manso. "Incentives and creativity: evidence from the academic life sciences." *The RAND Journal of Economics* 42.3 (2011): 527-554.

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.

Guinan, Eva, Kevin J. Boudreau, and Karim R. Lakhani. "Experiments in open innovation at Harvard Medical School." *MIT Sloan Manage. Rev* 54.3 (2013): 45-52.

Heckman, James J., and Edward Leamer, eds. *Handbook of econometrics*. Vol. 6. Elsevier, 2007.

Holmstrom, Bengt. "Agency costs and innovation." *Journal of Economic Behavior & Organization* 12.3 (1989): 305-327.

Imbens, Guido M., and Jeffrey M. Wooldridge. *Recent developments in the econometrics of program evaluation*. No. w14251. National Bureau of Economic Research, 2008.

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.

Kuhn, Thomas S. *The structure of scientific revolutions*. University of Chicago press, 2012.

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.

Lazear, Edward P. "The future of personnel economics." *The Economic Journal* 110.467 (2000): 611-639.

Lazear, Edward P., Ulrike Malmendier, and Roberto A. Weber. "Sorting in experiments with application to social preferences." *American Economic Journal: Applied Economics* 4.1 (2012): 136-163.

Lee, C. J. A Kuhnian critique of psychometric research on peer review. *Philosophy of Science*, 79.5 (2012), 859-870.

Lee, C. J., Sugimoto, C. R., Zhang, G., & Cronin, B. Bias in peer review. *Journal of the American Society for Information Science and Technology*, 64.1(2013), 2-17.

Lerner, Josh, and Julie Wulf. "Innovation and incentives: Evidence from corporate R&D." *The Review of Economics and Statistics* 89.4 (2007): 634-644.

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.

Manso, Gustavo. "Motivating innovation." *The Journal of Finance* 66.5 (2011): 1823-1860.

Milgrom, Paul, and John Roberts. "The economics of modern manufacturing: Technology, strategy, and organization." *The American Economic Review* (1990): 511-528.

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.

Nelson, Richard R. *The economics of parallel R and D efforts: a sequential-decision analysis*. RAND, 1959.

Nelson, Richard R., and G. Sidney. *An evolutionary theory of economic change*. Cambridge: Belknap (2005).

Romer, Paul. "Endogenous Technological Change." *Journal of Political Economy* (1990).

Rosenberg, Nathan. *Inside the black box: technology and economics*. Cambridge University Press, 1982.

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

Sauermann, Henry, and Wesley M. Cohen. "What makes them tick? Employee motives and firm innovation." *Management Science* 56.12 (2010): 2134-2153.

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

Schumpeter, Joseph Alois. *The theory of economic development: An inquiry into profits, capital, credit, interest, and the business cycle*. Vol. 55. Transaction publishers, 1934.

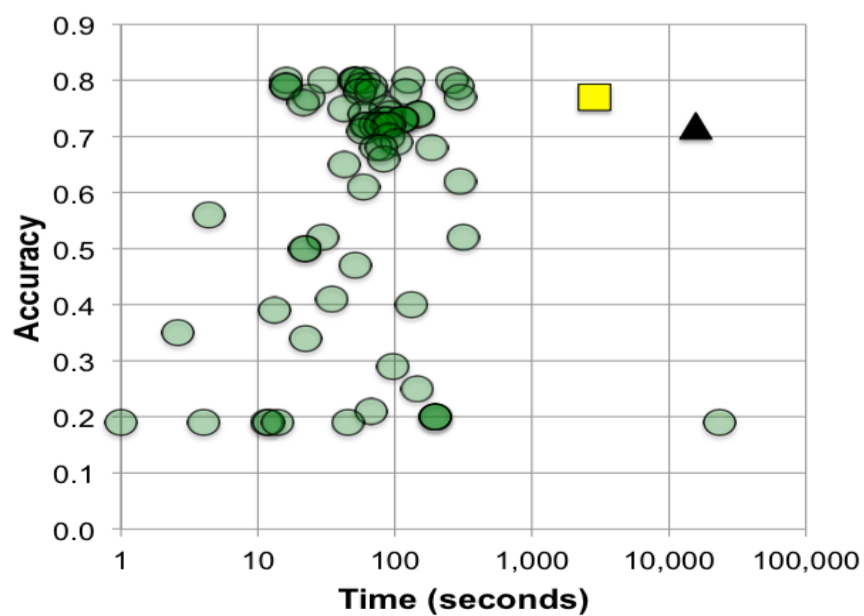
Stern, Scott. "Do scientists pay to be scientists?." *Management science* 50.6 (2004): 835-853.

Weitzman, Martin L. "Recombinant growth." *Quarterly Journal of Economics*(1998): 331-360.

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

Figures

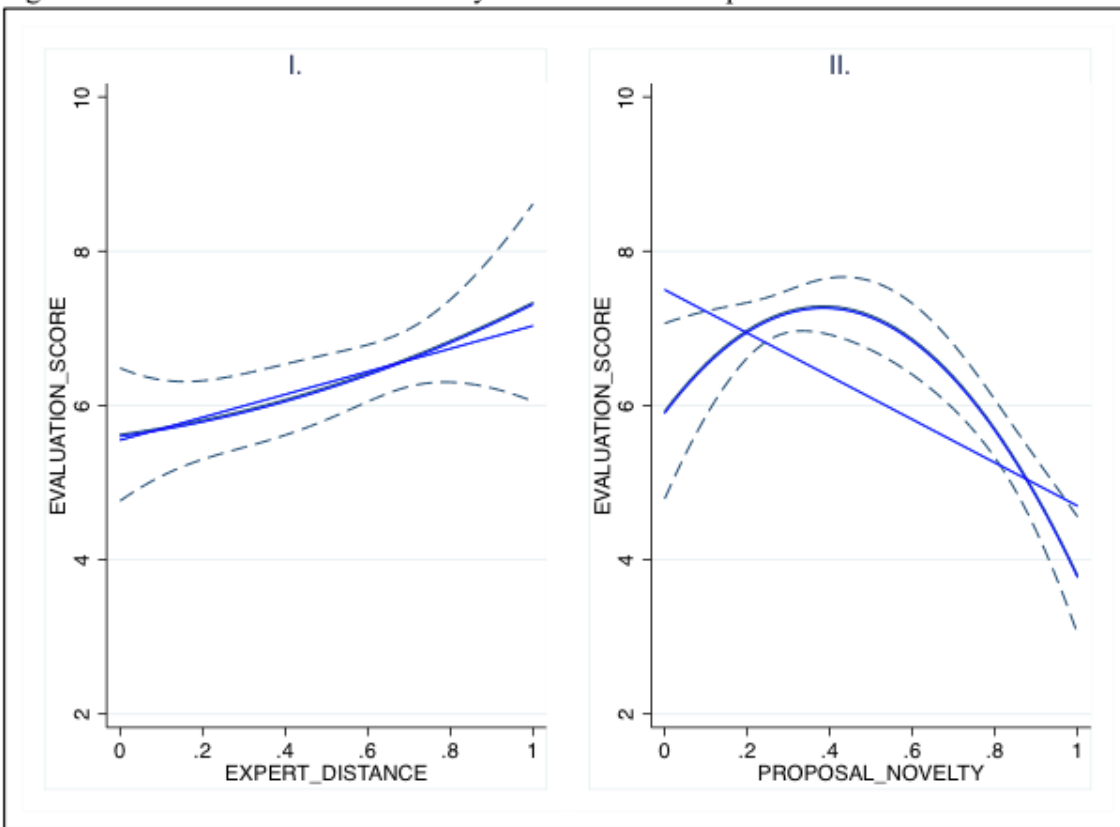
Figure 1 Accuracy score vs. speed of contest-commissioned immunoglobulin sequence annotation code



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.

Source: Boudreau et al. forthcoming

Tables

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	NASA	\$25,000	15 (255)	14	New algorithm to detect propeller objects in the rings of Saturn.

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	<i>De novo</i> 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	<i>De novo</i> 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.

Table 3 – Summary of Key Design Challenges and Solutions in Innovation Experiment Case Studies

Key Experimental Design Challenges	Study 1: Openness and Knowledge Disclosure (Lakhani et al. 2013, Boudreau and Lakhani 2015)	Study 2: Non-random Sorting in Innovation Contest Platforms (Boudreau and Lakhani 2011)	Study 3: Intellectual Distance and Novelty in Scientific Evaluations (Boudreau et al. forthcoming)
<i>Mechanisms</i>	- Comparative performance of innovation contests versus internal and grant-based research - The role of knowledge disclosure policy in incentives to innovate and the nature of the innovative search process	Preference for autonomous versus team work drives effort and performance in innovation contests while accounting for skills and pecuniary prizes	The role of intellectual distance and the recombinatorial nature of knowledge in systematically affecting the evaluation of scientific ideas while controlling for other forms of ad hominem biases
<i>Unit of Analysis/Replication/Sample Size</i>	- Population level technological performance and effort measures - Individual performance achievements - Knowledge sharing at population level - Recombinatorial analysis of knowledge approaches used to create software solutions - Limit replication to study population level policy effects by increasing sample size in large treatment groups	- Individual and team level analysis - Individual and team level effort and performance measures - Allow for multiple replications of treatment and controls – encourage broad entry	- Evaluation scores of research proposals as key driver - Measure distance between proposal and evaluators - Measure novelty based on all publications in PubMed - Assure multiple evaluators per proposal and multiple proposals evaluate by each evaluator (150 proposals, 142 evaluators Each proposal gets 15 evaluations)
<i>Selection/Treatment Effects</i>	Population level disclosure treatment allows for interpreting non-response as an incentive effect	- Design study as a “selection” experiment - Create treatments that allow for selection and matched pair allocation to autonomous versus team settings	Randomly allocate proposals to evaluators to derive variation in intellectual distance between evaluator and proposals
<i>Innovation Institution Treatments</i>	Knowledge disclosure bundles informational and motivational effects	Allow autonomous and team production	- Replicate NIH and HMS evaluation metrics while ensuring that evaluators do not contaminate each other by sharing evaluations - Separate out research hypothesis/idea from

			proposal feasibility
<i>Representativeness</i>	▪ Innovation problem solved of key interest to biomedical research community ▪ Quality scoring of solutions based on life sciences research objectives ▪ Solvers have background used by field organizations. Skills representative of researchers in field ▪ Prizes (\$6000) in line with platform specific rewards	▪ Choose NASA problem with clear and objective performance goals ▪ Instrument scoring system to enable objective measurement ▪ Offer large prizes (\$25,000 pool) to drive participation to enable replication and large sample size ▪ Subject skills match profile of programmers typically used to solve problems in organizations	▪ Solicit proposal from university researchers ▪ Recruit evaluators from medical school faculty ▪ Follow standard evaluation metrics ▪ Offer significant rewards for proposal success (\$25,000 combined reward pool)