

**Cumulative Innovation &
Open Disclosure of
Intermediate Results:
Evidence from a Policy
Experiment in Bioinformatics**

**Kevin J. Boudreau
Karim R. Lakhani**

Working Paper

14-002

July 2, 2013

Copyright © 2013 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.

CUMULATIVE INNOVATION & OPEN DISCLOSURE OF INTERMEDIATE
RESULTS: EVIDENCE FROM A POLICY EXPERIMENT IN BIOINFORMATICS

Kevin J. Boudreau and Karim R. Lakhani*

Abstract

Recent calls for greater openness in our private and public innovation systems have particularly urged for more open disclosure and granting of access to intermediate works—early results, algorithms, materials, data and techniques—with the goals of enhancing overall research and development productivity and enhancing cumulative innovation. To make progress towards understanding implications of such policy changes we devised a large-scale field experiment in which 733 subjects were divided into matched independent subgroups to address a bioinformatics problem under either a regime of open disclosure of intermediate results or, alternatively, one of closed secrecy around intermediate solutions. We observe the cumulative innovation process in each regime with fine-grained measures and are able to derive inferences with a series of cross-sectional comparisons. Open disclosures led to lower participation and lower effort but nonetheless led to higher average problem-solving performance by concentrating these lesser efforts on the most performant technical approaches. Closed secrecy produced higher participation and higher effort, while producing less correlated choices of technical approaches that participants pursued, resulting in greater individual and collective experimentation and greater dispersion of performance. We discuss the implications of such changes to the ongoing theory, evidence and policy considerations with regards to cumulative innovation. (*JEL O3, JO, D02*)

* Boudreau: London Business School and Institute of Quantitative Social Science at Harvard University, Regent's Park, London, U.K. NW1 4SA, fax: +44 (0)20 7000 8701, telephone: +44 (0)20 7000 8455, e-mail: kboudreau@london.edu; Lakhani: Harvard Business School and Institute of Quantitative Social Science at Harvard University, email: klakhani@hbs.edu. We are grateful to members of the Harvard Medical School communities for their contribution of considerable attention and resources to this project, including Ramy Arnaout, Eva Guinan and Lee Nadler. We also thank managers at TopCoder, including Jack Hughes, Rob Hughes, Mike Lydon, and Ira Heffan, who provided invaluable assistance in carrying out all aspects of the experiment and in designing and implementing the experimental platform. We thank expert computation and data scientists Po-Ru Loh, Hernán Amiune and Xiaoshi Lu for careful technical evaluations of the data algorithms developed within the experiment. We would like to thank several people for their comments, including: Lee Branstetter, Wesley Cohen, Carliss Baldwin, Erik Brynjolfsson, Chaim Ferschtman, Rebecca Henderson, Nicola Lacetera, Alan McCormack, Petra Moser, Ramana Nanda, Richard Nelson, Catherine Tucker and seminar participants at London Business School, Tel Aviv University, the National Bureau of Economic Research (NBER), and the Roundtable for Engineering Entrepreneurship Research (REER). Onal Vural provided excellent research assistance. All errors are our own. Boudreau would like to acknowledge financial support from a London Business School Research and Materials Development Grant and the University of Toronto, Rotman School of Management. Lakhani would like to acknowledge the financial support of the HBS Division of Research and Faculty Development. A Google Faculty Research Grant supported both authors.

I. Introduction

The Human Genome Project (HGP), which over a thirteen-year period identified more than 20,000 genes and sequenced the 3 billion chemical base pairs that make up the human DNA, has been described as one of the most ambitious large-scale scientific efforts in modern times (Watson 1990; Collins 2010). The project marshaled the work of over a thousand research scientists from more than 30 research laboratories spanning at least 19 countries. A notable feature of HGP’s governance was the implementation of the “Bermuda Principles,” whereby all participants released all their sequence data within 24 hours of discovery into the public domain (Contreras, 2011). This resulted in a near instantaneous disclosure of results with the intent that other investigators would build on these results and also to achieve better coordination among the researchers (Bently, 1996), with the hope of more rapidly advancing the science (Marshall, 1996; Cook-Degan and McCormack, 2001). This type of early disclosure of advances significantly departed from the usual practice of releasing final results and analysis in tandem with publishing a scientific journal article.

The practice of opening intermediate works—disclosing and granting access to them for reuse by others—has been observed in many areas and has long historical roots. Allen (1983), for example, showed that the mid-nineteenth century innovation process underlying the critical blast furnace technology for iron making, in UK’s Cleveland district, occurred through informal disclosures and formal publications of design and cost information between existing and new firms where technical and efficiency advances made by one firm were used to achieve further performance gains by others. Nulovari (2004) documents similar disclosure practices in the development of Cornish pumping engines, Bessemer steel and large-scale silk production. There are also numerous modern examples. In open source software projects, all manner of software development instructions are instantly made available for others to see and reuse, as developers make submissions to the code base (Lerner and Tirole 2005; Lerner and Schankerman 2011). Similar intermediate open disclosure practices have also been implemented in the Polymath Project for creating mathematical proofs (Growers and

Nielsen, 2011), Wikipedia (Greenstein and Zhu, 2012), computer hardware (Osterloh and Rota 2007), synthetic biology (Torrance 2010) and in Netflix’s \$1MM prize to improve its user movie rating prediction algorithm. Apart from these existing examples, there have been growing calls in recent decades for still greater openness, disclosures and access to intermediate works—early results, algorithms, programmes, technology platforms, materials, data, and techniques—upon which subsequent innovators, researchers, and inventors might build (Nelkin, 1982; Lin 2012; Royal Society, 2012).

To study the effects of intermediate disclosures on innovation, we devised an experimental approach. A total of 733 subjects (comprised of mathematicians, software developers, scientists, and data scientists) participated in a two-week experiment in which subjects developed and optimized a genomics analysis algorithm to solve a core computational biology problem faced both industrial and academic labs. The context can therefore be thought of as a “quality ladder” of sorts, with participants developing solutions in relation to a single, yardstick of quality and performance (Scotchmer, 2004, p. 132). Thus, we create a context in which a set of players interact in a common domain over time, as the state of the art evolves (with some players dropping out, some coming in, and some staying "in the game"). The experimental design essentially involved comparing independent subgroups of subjects, matched on ability and randomized on other characteristics, working on the same problem at the same time. Subjects also worked under similar institutional regimes in most every respect, except some worked under open regimes in which intermediate solutions were disclosed and made accessible to others within the same independent subgroup. Those in the closed control group, to whom those in the open regime were compared, worked under conditions in which intermediate solutions were not disclosed; only final solutions were disclosed at the end of the experiment. We observed characteristics of the individuals, including their ability levels, and precise measures of quality of solutions. We devised novel means of codifying the technical design approach used in each of the 654 intermediate and final solutions submitted during the course of the experiment.

This experimental strategy gives us a chance to study typically large-scale institutional design questions in a controlled context in which usual counterfactual outcomes can be observed. The problem addressed has itself been exposed to a past cumulative innovation process by both academic and industrial scientists over past decades of genome mapping (Altschul et al., 1990), while necessarily drawing on knowledge in computer programming and data science by the nature of the problem. The subjects are a mixture of professionals and students at various levels of academic and professional attainment who largely reflect the wide and diverse pool of representing a mix of skills in computer science, data science, and algorithm design skills relevant to this problem. The best solutions developed in the context of this experiment are indeed highly useful and have been disclosed to both the academic and industrial scientific communities for their use, on account of them exceeding the performance of benchmark solutions(see Lakhani et al., 2013). The experiment can itself be considered large-scale in exposing the problem to hundreds of subjects within each independent subgroup.

We implemented the experiment on an online platform so as to enable the application of precise treatments and controls and to allow fine-grained observations at the level of individual subjects and individual solutions contributed to the cumulative innovation process. Especially notable here is the ability to exploit precise measures of ability, effort, and activity and to discern technical approaches within individual intermediate solutions. Departing to some degree from a fully naturally occurring context also provides a rare opportunity to observe the complex series of endogenous processes as populations are exposed to distinct “rules of the game.” At least as important, an experimental design allows us to study completely independent “risk sets” of prospective innovators exposed to alternative institutional regimes. We can then derive inferences by making cross-sectional comparisons.

The reason for carrying out this research is that it is not clear what the effects should be. On the one hand, open disclosures of intermediate outputs might plausibly facilitate ex post reuse of intermediate works (Furman and Stern 2011; Williams 2013), propelling

cumulative innovation (Romer 1990). Intermediate disclosures of a wider set of inputs to the innovation process might also “lower the bar” for a more diverse set of entrants to participate (Murray et al., 2009). Further, high incentives of innovators might plausibly be maintained, depending on what downstream innovators do with the disclosed material and if innovators can somehow be recognized and rewarded for the disclosures they make. In a sense, this is similar to disclosure in say academic publication (Dasgupta and David 1994; Stephan, 1996), the patent system (Kitch 1977) and subsequent to prize contests (Sobel, 1996); however, open disclosures of intermediate works often mean stepping way from these established institutions (ex: a completed article, product, or invention) to facilitate a much freer flow of knowledge (e.g., Furman and Stern, 2011; Williams 2013), raising again the question of rewards, recognition and incentives. The implementation of an intermediate disclosure policy may be in conflict with a basic tenet of the economics of innovation that diminished control and appropriability over knowledge assets also has the potential to reduce incentives to make costly investments, exert high effort in their creation—or even to participate in the innovation process altogether (Nelson 1959; Arrow, 1962). Open disclosures of intermediate work might also affect choices of technical approaches—on the one hand, perhaps inducing differentiated innovation pathways (Murray et al., 2009) and wider range of possibilities considered (Weitzman 1998). Greater disclosure might also expand the range and types of actors who can participate effectively in the innovation process (Murray et al., 2009). And in so doing it may expand the variety of approaches taken. Alternatively, more disclosures may just steering innovators towards incremental accretion of improvements along existing pathways (Acemoglu, 2012).¹

We organize our analysis around investigating effects on both (i) the rate of innovation and performance attained and (ii) the range and nature of technical approaches assayed. Given myriad possible dynamics that might be set into motion, we do not constrain our analysis or impose a structural interpretation. Rather our analytical approach is to begin by

¹A distinct literature considers strategic voluntary disclosures (ex: Haeussler et al., 2009).

documenting differences in (i) and (ii) and then to explain these differences in terms of the “types” of innovators choosing to enter and participate, the level of effort they exert, and the particular technical choices taken.

We find a range differences produced by disclosures of intermediate works, beginning with reductions in participation and activity levels on the order of 30% or more, depending on the measure. Effects are observed across the entire distribution of higher- and lower-ability subjects. We interpret these patterns as consistent with diminished economic incentives in a system of freer disclosures, particularly where disclosed knowledge was reapplied by downstream innovators to addressing the same problem addressed by upstream innovators. While there might be ways to attenuate this drop, the consistency of these results with economic theory and remarkably large magnitude of the drop observed here deserve emphasis.

Participants in the open regime chose to focus their efforts, lower as they were, on most performant established technical approaches. This led participants in the open regime to achieve higher performance scores than those in the closed regime. Scores were disproportionately clustered in the “upper tail,” reflecting a concentration of solutions on most performant technical approaches. The Herfindahl index for technical approaches, for example, was 52% higher in the open regime. Thus “learning” here was constituted of both knowledge transmissions and responding to signals conveyed in the open regime around the quality of other solutions and actions of others. This led to coordinated technical choices, focusing improvements along existing innovation pathways.

By contrast, participants in the closed regime engaged in substantially higher levels of experimentation, with each active participant in the closed regime assaying 17% more technical approaches, as measured by combinations of techniques assayed. The difference in aggregate experimentation was even higher—42% greater number of unique combinations of techniques assayed—resulting from the combination of higher numbers of participants, higher levels of individual experimentation, and greater independence of choices of technical approaches pursued. As each of the aforementioned effects applied relatively uniformly across

the skill distribution, implications apply equality to average outcomes as they do to the maximal outcomes in each comparison group.

While heterogeneity of participants contributed to the range and level of results, open disclosures of intermediate results did not change the “types” of participants who chose to enter and actively participate, only their number. Despite considerable heterogeneity in the overall pool of subjects (including 69 countries represented, a range of ability levels and divergent technical interests), the distribution of skills and other characteristics of those choosing to participate were remarkably similar across the regimes.

In relation to the academic publishing system, the patent system or prizes with disclosures of winning solutions—which attempt to reconcile and moderate tensions between upstream or *ex ante* incentives, and downstream or *ex post* learning and reuse—the patterns observed here indicate that intermediate disclosures strike more extreme tradeoffs. In a nutshell, within this single-problem environment, intermediate disclosures favored exploitation of existing knowledge, rather than exploration across other potential technical approaches. Intermediate disclosures also discouraged effort levels taken by any one individual, including highest-ability experts. Intermediate disclosures were found to be effective in the context studied here. However, intermediate disclosures should, for the same reasons, possess inherent disadvantages in contexts in which there are high returns to exploring unknown approaches or in providing high incentives to most expert innovators.

Our findings contribute to a recent stream of research on the design of institutions supporting cumulative innovation, and particularly the effects of open disclosures and access (e.g., Aghion, Dewatripont and Stein 2008; Mukherjee and Stern 2009; Murray et al., 2009; Gans and Murray 2012). We underline the importance of this work, finding evidence that the design of disclosure policies can produce first-order changes in the innovation process. We deviate somewhat from this work with our focus on intermediate disclosures (rather than, as in this literature, comparisons of distinct disclosure systems as academic publication and

patenting²). Whereas empirical research on open disclosures and access (e.g., Murray and Stern, 2007; Rysman and Simcoe 2008; Murray et al., 2009; Boudreau 2010; Furman and Stern, 2011; Galasso and Schankerman; Williams, 2013) has focused on effects of disclosures on *ex post* reuse, we study factors affecting advances in technical performance. Williams (2013) is closest to our interest in intermediate disclosures, studying genetic sequences placed in the public domain, as part of the HGP. Our finding of intermediate disclosures discouraging diversity and exploration in this single-problem environment might also be contrasted with Murray et al.’s, (2005) finding of evidence suggesting open disclosures encouraged greater reuse of a disclosed technology across alternative problems and applications.

Our work also complements the pioneering contributions attempting to consider the endogenous emergence of diversity and experimentation (e.g., Murray, et al. 2009; Acemoglu, 2012), themselves based on a long line of rich theorizing on the rate and direction of scientific and technical advance (e.g., NBER, 1962; Rosenberg, 1976; Mokyr 2002; Lerner and Stern, 2012). Above all, we find that lack of coordination—particularly in context of higher incentives and activity—that leads to heightened diversity in the closed regime by maintaining independence of experimentation choices (Nelson, 1961). By contrast, we find no evidence within this single-problem environment that open disclosures widened the innovation process to more diverse actors (Murray, et al. 2009), nor produced a wider consideration of possibilities (1998); rather, open disclosures contoured efforts towards less diversity (Acemoglu, 2012). In addressing the disclosure of intermediate works, in particular, this paper also links distinct research in areas such as open source software (e.g., Lerner and Tirole, 2005), digital “mashups” of music and multimedia (Lessig 2009) and other seemingly exotic corners of the modern economy to the study of more mainstream institutions. As regards research on cumulative innovation and growth, more broadly (e.g., Romer, 1990, Aghion and Howitt, 1992, Green and Scotchmer, 1995; Aghion et al. 2005; etc.) we illustrate the possibility that

²Nor is our intent here one of contributing to the debate on the workings and effectiveness of patents (e.g., Moser 2013)—which involves questions of the effectiveness of disclosures, apart from the question of the effects of disclosure, in principle..

“lab”-like settings can be exploited to study the innovative process.

The paper is structured as follows: Section II presents the experimental design. Section III describes the data. Section IV reports results. Section V discusses concludes.

II. Experimental Design and Methods

A. *The Problem Addressed by Subjects in the Experiment*

We anchored our experimental design on a problem for which we expected to observe a sequence of cumulative innovation. We worked closely with colleagues at the Harvard Medical School to “layer” our experimental treatments within a biomedical research effort to create solutions to a bioinformatics algorithm development problem (see Lakhani et al., 2013 for greater detail). The specific problem involved creating new algorithms enabling the identification and annotation of the constituent gene components of immune system related genetic sequences that had been recombined and mutated. Subjects in our experiment had to develop *de novo* operational algorithms, written in computer code, that could annotate 10^5 genetic sequences (a typical high throughput genetic sequencing run) with performance characteristics (accuracy of annotation and speed of annotation) that would meet or exceed current benchmarks used by academic and industrial research labs such as the US National Center for Biotechnology Information’s MegaBLAST program and the internal Harvard solution.

A number of features of this problem made it attractive for studying processes of cumulative innovation, where both continuous advance and experimentation might play a role. As a problem sitting at the intersection of software development, mathematics, computer science, and bioinformatics, it is nontrivial and challenging, it is subject to the application of a range of techniques and draw from a range of knowledge. Further, as an optimization problem without a closed-ended solution, but rather one in which numerical methods could be applied to eke out incremental gains, we could observe how efforts taken by subjects result

in additional gains (rather than simply observing discrete outcomes of correct or incorrect outcomes). In these senses of representing a complex, data-intensive numerical optimization problem that cuts across knowledge domains, it is a kind of problem that might be encountered in a wide range of contexts, from industrial innovation to academic research. Focusing on algorithm development, we are also able to treat intermediate solutions themselves as a primary input to subsequent innovation processes—to which other subjects might practically gain access through disclosure. (Conducting an analogous experiment in which intermediate work took the form of physical or living materials, such as research mice, would make carrying out the experiment far less practicable for obvious reasons.) Working in digital format, where solutions are codified in computer instructions, carries a range of other advantages that are detailed in following discussion.

B. The Experimental Context

We ran the experiment on an online platform to practically and precisely implement treatment and controls, to enable the observation of fine-grained measures, and to draw from a wide, relevant, and diverse population of prospective participants. The experiment was run on the TopCoder platform—a large online platform on which algorithmic developers, data scientists, software developers, mathematicians, and practitioners of these areas within various industries participate in a regular stream of competitions to solve problems. Among the several hundred thousand expert algorithm and software developers who are members of the platform, 733 chose to participate in the experiment. In recruiting subjects through the TopCoder website and distribution lists, we explained the duration of the experiment, the magnitude of cash payoffs, and that the problem would be an algorithmic problem for the Harvard Medical School. Among the 733 individuals who participated in this experiment, roughly half (44%) were computer and data science professionals and the remaining were students at various levels of achievement. Participants came from 69 countries.

Running the experiment on the TopCoder platform provided a medium in which we could implement a precisely controlled environment. We worked with TopCoder executives and

engineers to implement the platform designs described herein. A number of features of the environment were common to all experimental groups. For example, all development and interactions were to take place on the online platform. Working through a web-based Internet connection, participants were given the problem at the start of the experiment. It was also stated at this time that scoring would be based on a combination of accuracy and speed of execution. The web-based interface featured a development area in which participants could develop their algorithmic instructions in computer code. They could then compile and test that code in a wide range of leading computer languages, submitting their code for compilation to the central platform where it was run and—if successfully compiling—would then be tested against test data sets that subjected solutions to a raft of tests. A numerical performance score related to accuracy and speed of execution would then be returned to the participant. Thus, it was not possible to receive direct feedback on the quality of the submission “off line.” Participants’ final submissions were taken as the basis for determining winners.

C. The Experimental Institutional Regimes

The experimental design involved comparing matched independent subgroups of subjects, working on the same problem, at the same time, in with same general features of the institutional context—but under distinct disclosure regimes. In our open regime, all intermediate solutions were available for review by all other participants within the regime. In implementing this regime, there is of course no single “disclosure lever” to implement open disclosure of intermediate solutions (cf. Dasgupta and David, 1994). A set of closely related design changes were devised to implement this institutional design. This design begins with the choice to implement a relatively frictionless “click-through” system to allow participants to instantaneously observe each others’ solutions through the same web interface on which they conduct their development. Further, we elected for the disclosure of intermediate solutions to take the form of the working computer code, itself. Consistent with open disclosure systems

in practice, we implemented an accompanying system of attribution whereby those reusing ideas could “cite” the solutions on which they drew. (This would also have implication for payoffs, as below.)

Our comparison or “control” group was a closed regime, which involved neither disclosure nor access to intermediate solutions over the two-week experiment. Only final solutions were disclosed after the conclusion of the experiment. No technical mechanism was made available for sharing within the experimental context during the experiment and participants were explicitly instructed not to share solutions under threat of being disqualified. Nonetheless, multiple intermediate solutions were submitted to the platform prior to final solutions on account of the need to receive feedback on solutions in a regular trial-and-error development process.

For both the groups working under rules of open disclosure and those working under rules of closed nondisclosure, payoffs were awarded after each week. The top five participants in each group were publicly acknowledged on the TopCoder website. The top five slots in each of the two weeks were allocated a total of \$1000 in cash prizes (\$500, \$250, \$125, \$75, and \$50). In Open/Disclosure, half of each of the monetary prizes was allocated to those cited in winning solutions. Those cited within the winning solutions were also publicly acknowledged on the TopCoder website, alongside winners.

D. The Assignment of Subjects to Independent Trials

One of the most important design decisions related to the size (equivalently, number) of trials and comparison groups constructed with the 733 experimental participants. Above all, we prioritized creating groups that were as large as possible. In so doing, we could best reflect “policy experiment” in which populations of prospective entrants would be exposed to alternative regimes and could elect to enter and actively participate, or not. Thus we created a single large “Open/Disclosure” group that engaged in development over the two weeks under a regime of open disclosures of intermediate work; and we created a single large

“Closed/Nondisclosure” group, which engaged in development over the two weeks under a regime of closed nondisclosure. Just one other group was constructed to supplement our two main Open/Disclosure and Closed/Nondisclosure groups. Given our priority to construct large trials and consequently run a minimum of trials, the supplementary regime was constructed to gain greater assurance that results in main groups were not somehow eccentric, while also providing another perspective to support interpretation of patterns. Therefore, we constructed a “Mixed” group. The Mixed group worked under a closed regime during the first week of the experiment and an open regime during the second week. In each of the three groups, participants were matched on skills otherwise randomly assigned. This was done by ordering the set of subjects according to their TopCoder skill rating (Section III) from top to bottom, and then assigning successive sets of ability “triplets” to each of the three groups.

E. The Approach to Statistical Inference

In broadest outlines, our empirical strategy is analogous to that of past research, involving comparisons across regimes. Rather than attempting to directly discern the particulars of the endogenous cascade of interacting dynamics within an on-going cumulative innovation process, prior research has focused on making more or less square comparisons between outcomes in open regimes and comparison regimes. Predominantly, these comparisons have been made in the style of differences-in-differences, as regimes have switched while others have not (e.g., Huang and Murray, 2009; Murray et al., 2009; Boudreau, 2010; Williams, 2013). Closely kindred papers making comparisons across academic and industrial teams, both making disclosures through similar mechanisms, has proceeded with analogous cross-sectional comparisons, attempting to more or less match or control for the nature of the discovery (e.g., Moon, 2011; Bikard, 2012). Here too we make square comparisons across regimes, in this case by precisely constructing matched comparison groups.

Given the design, we are able to regress a series of outcome variables, denoted here simply as y , on an indicator variable switched to one for participating in Open/Disclosure

(*OpenDisclosure*) and a constant, α , where β is a coefficient and ε is a zero-mean error term, with subjects indexed by i :

$$y_i = \alpha + \beta \cdot \textit{OpenDisclosure}_i + \varepsilon_i \quad (1)$$

Apart from cross-regime variation, it should be noted too that subject characteristics denoted by θ —and particularly ability skills level on which groups are precisely matched—serve as another source of variation we are able to exploit to derive inferences. Both the mean outcome and response to open disclosure plausibly depend on the type of participant (e.g., Murray et al., 2009). The following expression captures this elaboration of the analysis, where the zero-mean error term, ε , has been re-defined accordingly:

$$y_i = \alpha(\theta_i) + \beta(\theta_i) \cdot \textit{OpenDisclosure}_i + \varepsilon_i \quad (2)$$

We estimate all models using both OLS and nonparametric methods, as elaborated within the analysis. To emphasize, the design allows us to compare matched sets of *prospective* entrants, as we wish to observe both participation and non-participation stimulated by the experimental regimes. Nonetheless, we also report comparisons of outcome variables, conditional on participation. In the very least, such comparisons remain substantive and meaningful descriptive comparisons. Whether differences in outcomes conditional on participating can be interpreted as reflecting a treatment or selection effect are questions explored within the analysis. These questions relate closely to theoretical claims that greater openness may “lower the bar” and bring greater diversity to the pool of entrants (e.g., Murray et al., 2009) or perhaps attract participants of varying quality (cf. Furman and Stern, 2011).

III. Measurement and Data

We collected information on the characteristics and activities of subjects within the experiment, and characteristics of solutions they generated. Our econometric analysis focuses on cross-regime comparisons concerning the 245 subjects assigned to Open/Disclosure and a comparison group of 244 subjects assigned to Closed/Nondisclosure. Within the results, we also present data from the 244 subjects assigned to the Mixed regime, in a supplementary fashion. This section presents the data set and variables collected during the experiment.

A. Measuring Performance

As regards quality, here we benefitted from an automated scoring suite, made possible by running the experiment on an online platform, whereby subjects could submit their code to the platform and nearly instantaneously receive feedback on the quality of their solution as it was tested against a test data set that resided on the platform. The scoring mechanisms, algorithm, and test data sets were developed by personnel from Harvard Medical School and TopCoder working in collaboration. A numerical score was recorded in each instance, reflecting the speed and accuracy of each submission, as applied to the test data sets. The final score of each subject, *ProblemSolvingScore*, was that of the final submission of each participant. It was on the basis of this score that final rank order was determined. The variable *ProblemSolvingScore* was bounded from below by a minimum score of -7.14 and ranged upwards to a maximum of 0.80, where larger values indicate higher performance.³ Those subjects choosing not to participate were given the minimum score.

B. Measuring Technical Approaches

³There are no substantive interpretations of negative versus positive scores. An original raw scalar (x) ranging upwards from zero to 1255.5 was transformed as $-\ln(1255.6-x)$, to equally weight what might otherwise appear to be small differences near the top of the linear scoring spectrum. Therefore, the value -7.14 is simply $-\ln(1255.5)$. This transformation does not qualitatively change results, but increases statistical significance.

Apart from developing a precise measure of solution quality, we observed technical approaches taken. This too was made possible by running the experiment on an online platform, so as to observe solutions. We hired three Ph.D.-level experts to examine each of the 654 intermediate and final solutions. These experts identified ten key elemental optimization techniques used within the population of solutions. (Details are described in Lakhani, et al. (2013)). Thus, each submission was coded in terms of a 10-digit binary code representing a combination of techniques. There were 56 unique combinations developed across the entire experiment. (The performance and quality of a given solution can be understood as depending on the technical approach or combination of techniques used and the degree of refinement of the approach.) Within our analysis, we study the number of techniques assayed across both intermediate and final solutions by a subject (*NumTechniquesTried*). We also study a separate count of the subset of those techniques that appear in the final solution of a subject (*NumTechniquesInFinal*). In addition, we analyze the number of distinct approaches or combinations of techniques that were assayed by each subject (*NumCombinationsTried*).

C. Measuring Participation, Activity and Effort

We also observe several measures of activity and effort exerted by subjects. At a most basic level, we measure whether individuals decided to actively enter and participate, with an indicator variable switched to one for those participants who submitted at least one solution (*Participated*). As submissions served as a means of engaging in trial-and-error development, the count of number of submissions made by a subject (*NumSubmissions*) also served as an indication of the level of development activity. We also separately asked active participants to self-report the number of hours they worked (*HourWorked*) in a questionnaire after the completion of the survey and prior to announcing winners, for which 60% of participants responded. Given the self-reported nature of these data and their partial coverage, we simply refer to these data as a means of corroborating our analysis of *NumSubmissions* as our preferred indication of activity levels.

D. *Treatment Effects and Subject Characteristics*

We record an indicator variable switched to one for subjects assigned to Open/Disclosure (*Open/Disclosure*). We also measure characteristics of individual subjects. Most important is a precisely-measured rating of their ability or skill level (*SkillRating*). The skill rating of participants on the TopCoder platform is based on an Elo-based system (Maas and Wagenmakers, 2005), that estimates skill on the basis of historical performance in similar algorithmic problem-solving exercises. (The average participant engaged in dozens of problems prior to the experiment.) The Elo system is standard in a range of contexts from chess grandmaster tournaments to US College Bowl systems to the National Scrabble Association and the European Go Federation. We use a normalized version of this measure (i.e., with mean set to zero and deviations measured in numbers of standard deviations). We supplement this ability measure with measures of other characteristics of individuals, including the technical area they were most interested in at the time they joined as a member to the TopCoder platform, chosen from a finite list of options presented to them at that time. We also observe country of origin, again based on self-reporting at the time they joined as a member to the TopCoder platform. (Descriptive statistics of these latter variables are provided within the analysis, itself.)

IV. Results

A. *Overview of Results*

Here we begin by presenting broad patterns of outcomes to establish most basic facts, with following subsections proceeding to analyze and explain these facts.

Regarding the advance in quality and performance of solutions, an overview of the streams of submissions generated in each experimental regime appears in Figure I. A total of 654 solution submissions were made over the two-week experiment across the three independent groups. Intermediate solutions appear as grey dots, indicating timing (x-axis) and quality

(y-axis). Final submissions are black dots. The graphs also trace maximal frontier lines and moving averages. These descriptive patterns, on their own, begin to suggest differences between the regimes. In our main comparison groups, the standard deviation of final scores in Open/Disclosure is 2.88 points and that in Closed/Disclosure is smaller, but still large, at 1.81. This is because scores are bounded from below and, within Open/Disclosure, solutions are clustered on the “right tail” of maximal scores. In Closed/Nondisclosure, solutions are clearly far more evenly dispersed from low- to high-quality solutions. Within the supplementary Mixed regime, we observe patterns that would appear to be somewhat consistent with these distinctions, with relatively evenly dispersed solutions in the first (closed) week and submissions becoming sparser and weighted towards the maximal frontier in the second half.

<FIGURE I>

Regarding the extent and range of experimentation that the experimental regimes lead to, Figure II plots the generation of unique combinations or “technical approaches” over time in each of the independent groups, where a unique approach is the first appearance of a submission embodying a particular combination of elemental techniques. We include here, once again, the supplementary Mixed regime to illustrate that it traces an intermediate level between Open/Disclosure and Closed/Nondisclosure. We find large differences, with 27 unique combinations assayed in Closed/Nondisclosure and 19 in Open/Disclosure: 30% fewer technical approaches assayed in Open/Disclosure than in Closed/Nondisclosure. Apart from the large magnitude of differences in levels of experimentation across combinations of techniques, differences between Open/Disclosure and Closed/Nondisclosure are systematic and stable in the sense of having no “crossing points” between the curves.

<FIGURE II>

Our main analysis to follow analyzes these differences and their causes, focusing on our main open and closed comparison groups. We return to the mixed regime as a means of validating results and our interpretation in summarizing and discussing patterns.

B. *Does Open/Disclosure Affect Incentives, Activity and Levels of Participation?*

Here we document evidence of much lower activity and incentives in Open/Disclosure. We begin by studying subjects' decisions to enter and actively participate (indicated by having submitted at least once). Results of OLS regressions with bootstrapped standard errors are presented in Table I. Model (1) simply regresses an indicator for having participated on a constant and an indicator variable for participating in Open/Disclosure. In Open/Disclosure, the probability of participating was 14%, or 5% lower than in Closed/Nondisclosure. Model (2) adds our skill measure as a control, and model (3) adds an interaction between skill and the indicator for Open/Disclosure. While the positive coefficient on *SkillRating* suggests higher-skilled individuals have a generally higher propensity to participate—whatever the regime—the interaction term is not significantly related to participation decisions. Therefore participation is generally lower in Open/Disclosure, with no difference detected across subjects at different skill levels.

To verify these results, Figure III presents the results of a more flexible model, comparable to model (3), but which allows model (3) to be re-estimated non-parametrically using a locally-weighted second-order polynomial employing an Epanechnikov kernel (DiNardo and Tobias, 2001; Silverman, 1986). Confidence intervals presented in Figure III are simply estimated parametrically so as to increase efficiency in estimating these bounds.⁴ This flexible estimator affirms the effect of open disclosure on entry and active participation appears to be invariant to skill level. Therefore, open disclosure did not result in a change of skills distribution, only a lower level of entry in general.

⁴Using parametric or nonparametric estimates of these bounds does not substantially affect the conclusions of the analysis presented in Figure III; however, efficiently estimating these bounds becomes more important in later analysis in which we use just a subset of the observations to compare just those subjects who chose to participate.

<TABLE I>

<FIGURE III>

Also consistent with lower incentives in Open/Disclosure, we observe lower activity and effort. Model (4) regresses our observational measure of activity levels, the number of submissions (an indication of trial-and-error activity), on an indicator variable for participating in Open/Disclosure and a constant term. The average number of solution submissions made per individual in Open/Disclosure was 0.4, 0.9 less than the average of 1.3 submissions in Closed/Nondisclosure. Adding our skill measure and interaction with Open/Disclosure in models (5) and (6), we see that numbers of submissions were higher in Closed/Nondisclosure and among higher skilled participants. The negative interaction term in model (6) is negative, indicating that—counting the non-participants who did not submit as zero—that the boost of activity in Closed/Nondisclosure is higher still among the higher-skilled subjects. Adding these additional regressors does not change the estimated coefficient on *OpenDisclosure*. These patterns are clarified in the first panel of Figure IV, which graphically shows the estimated relationships in the form of flexible, nonparametric estimates.

Models (7) through (9) re-estimate models (4) through (6), but just for participants who actively participated, i.e., with at least one submission ($N = 79$). Therefore, these regressions measure differences of activity among active participants. As in model (7), active participants in Open/Disclosure submitted 3.1 solutions—3.9 fewer than the 6.9 in Closed/Nondisclosure, as reported in model (7). Therefore, by this measure activity levels were less than half in Open/Disclosure. Measures of skills and interactions between skills and the Open/Disclosure indicator are insignificant, conditional on participating, as reported in models (8) and (9). Therefore, it appears the earlier effects of these variables were driven by the decision to participate (i.e., the change between zero and one submissions); the effect of Open/Disclosure on activity levels conditional on participating appears to be constant across the skills distribution. These patterns are affirmed in flexible, nonparametric estimates presented in the second

panel of Figure IV. Despite a wider confidence interval on account of fewer observations, the number of submissions is a good deal higher in Closed/Nondisclosure—and flat across ability levels. These earlier finding that likelihood of participation with Open/Disclosure does not interact with skill is equivalent to saying the skills distribution in either regime is the same. (Later, more exhaustive tests confirm this.) Therefore, we interpret estimated effects of Open/Disclosure in models (7) through (9) as treatment effects, whereby the regime influences behavior conditional on entering.

To further validate our interpretation of lower incentives and effort in Open/Disclosure, we analyzed a separate measure of effort and activity, the number of hours worked by participants. This provides a more direct measure of effort, but with the disadvantage of being self-reported. The survey achieved a response rate of 60% (N = 47). Of those active participants who responded, those in Open/Disclosure worked 15 hours on average, while those in the Closed/Nondisclosure worked 21 hours. As presented in the third panel of Figure IV, despite the partial response rate and self-reported data, we see large and statistically significant differences).⁵

<FIGURE IV>

C. How Does Open/Disclosure Affect the Type of Subjects Participating?

Here we examine whether there were differences in the types of individuals who entered and actively participated in open and closed regimes. As regards the quality or skill level of entrants, the preceding analysis found that the lower participation in Open/Disclosure was not related to skill, suggesting no systematic differences in the *distribution* of the skills variable, *SkillRating*. Beyond these earlier tests, the first panel in Figure V explicitly plots the empirical cumulative distributions of *SkillRating* in both Open/Disclosure and

⁵A linear OLS regression model finds that reporting participants in the Open/Disclosure work 6 fewer hours than those in Closed/Nondisclosure, significant at $p = 10\%$. The linear model does not detect a statistically significant relationship with *SkillRating*, despite the apparent downward slope in Figure IV. The interaction between skills and an indicator for the Open/Disclosure regime is also insignificant.

Closed/Nondisclosure and fitted cumulative Normal distribution functions. The plots clarify similarities in the curves. Tests for differences in first and second moments of these distributions are presented in Table II, finding no significant differences. Apart from these parametric tests, we also applied a non-parametric Kolmogorov-Smirnov test of differences in distribution (Justel et al., 1997), which essentially measures the supremum of the set of vertical distances between the empirical cumulative distribution functions. (Thus, this test makes no assumptions regarding the structure or form of distributions.) We are unable to reject the null hypothesis that the observed distributions are drawn from the same underlying distribution in relation to the maximal distance between the distributions. (In relation to Closed/Nondisclosure having lower skills than Open/Disclosure, at $p = 18\%$; in relation to Open/Disclosure having lower skills than Closed/Nondisclosure, at $p = 80\%$).

Supplementary to these tests on *SkillRating*, we performed analogous tests on other measures likely to be correlated with ability, including the number of instances in which they had participated in analogous TopCoder events to develop new algorithms in the past. We also studied the year in which they joined the platform. Empirical distributions are again plotted in Figure V and parametric comparisons performed in Table II. We again find remarkably similar distributions and no statistically significant differences.

<FIGURE V>

Apart from finding no evidence of differences in the distribution of “high” or “low” quality participation, it remains possible that other kinds of differences might still exist. For example, apart from general algorithmic problem-solving ability, active participants may differ in specialized knowledge or experience. We find no evidence of such differences. As regards technical interests, TopCoder collected on the stated main technical interest of participants from 45% of subjects. These include categories described as follows: Broadband; Data, Voice, Video Convergence; Game Software Development; Graphic Design; Handheld; Networking; Security; (General) Software Development; Web; and Wireless. General Software Develop-

ment, Games Development, and Web Development were the largest categories. We report differences across the most common categories and “Other” (including non-respondents) in Table III and find no evidence of significant differences. Even Herfindahl indices are remarkably similar. For Open/Disclosure the Herfindahl index for all individual categories, including non-responses is 0.396; for Closed/Nondisclosure it is 0.427. If we drop the non-respondents and calculate the Herfindahl just for participants who reported a main interest among the earlier listed categories, the Herfindahl drops considerably but remains similar across regimes, at 0.606 and 0.740.

Likewise, TopCoder collected country data on 100% of members. Table III also shows the distribution of the population of subjects and those of active participants in either regime, listing the top countries—India, USA, Russia, China—along with 65 “Other” countries. We again find no statistical differences. The Herfindahl for participants in Open/Disclosure is 0.082 whereas that for Closed/Nondisclosure is 0.089. Therefore, despite the considerable scope for differences to emerge across the many countries with their respective comparative advantages, local knowledge, and local labor market conditions, we find little difference.

<TABLE II>

<TABLE III>

D. *Does Open/Disclosure Affect Performance and Technical Approaches?*

It is not possible to detect significant differences in performance scores across all subjects (assigning minimum scores to non-participants), as in the first panel of Figure VI. This might itself be regarded as remarkable, given subjects in Open/Disclosure worked 6 fewer hours than did those in Closed/Nondisclosure, while having no observable skill differences. Moreover, the problem solving scores are even statistically and substantially higher in Open/Disclosure, once examining performance conditional on participating, as shown in

the second panel of Figure VI. Active participants in Open/Disclosure attained 0.67 higher scores on average (half a standard deviation of the score measure). Given the equal composition in observable characteristics in each regime, we interpret these differences as largely reflecting causal treatment effects rather than selection effects.

<FIGURE VI>

The earlier results concerning performance would suggest participants learned a great deal more within Open/Disclosure on the basis of observing others' solutions, rather than only engaging in private trial-and-error experimentation. It follows we should expect to see tangible differences in the nature of solutions and the solution process. These differences can indeed be readily observed when comparing the number of individual optimization techniques assayed across all submissions, as in panels 1 and 2 of Figure VII. Conditional on participating, those in Open/Disclosure tried out 0.81 more techniques on average. They also implemented 0.64 more techniques in their final solutions, as in panels 3 and 4 of Figure VII. Deploying greater numbers of optimization techniques in large part accounts for the higher performance attained by those in Open/Disclosure. (Note, however, that despite the importance of numbers of techniques in affecting performance score, the very highest scoring techniques did not employ the greatest number of techniques; they benefitted from the choice of particular combinations of techniques and the effectiveness with which these were refined and implemented.)

<FIGURE VII>

Despite greater learning in Open/Disclosure, there are indications of considerably lower levels of experimentation. This was first suggested by more clustered solutions documented in Figure I. It should also be noted that despite having tried 0.81 more techniques and

having been exposed to many more, participants in Open/Disclosure assayed fewer combinations of these techniques (panels 5 and 6 in Figure VII).⁶ Conditional on submitting, those in the Open/Regime assayed .71 fewer combinations of techniques, with 1.67 versus 1.96 combinations assayed in open and closed regimes. This is especially notable given that the greater number of techniques individuals could “see” in Open/Disclosure would on their own allow for greater number of combinatorial possibilities. In addition to lower overall levels of experimentation (Figure II) and lower levels of individual experimentation (Panels 5 and 6 of Figure VII), there are also no more novel techniques developed *per individual*. While differences are not statistically significant, it is notable that each subject in the open regime even develops 0.03 *fewer* novel combinations on average than those in the closed regime and each active participant develops 0.02 fewer, if we define novel techniques as the first appearance of a given combination within a regime. This is because developing novel, unprecedented combinations within Open/Disclosure should be inherently easier where there are fewer unique combinations altogether.

The lower levels of experimentation by participants in Open/Disclosure also appear to be more targeted towards performant established solutions. This too could first be appreciated from Figure I with progressively greater clustering of solutions in the upper tail around maximal scores. This convergence itself accounts for the significant differences in performance in Open/Disclosure. The simple fact that higher performance was achieved with lower levels of trial-and-error (i.e., submissions) and with narrower experimentation (i.e., combinations of techniques) itself demonstrates more coordination and “targetedness” of contributions in Open/Disclosure. We performed several additional comparisons to further corroborate these points. For example, we rank ordered the 56 distinct combinations of techniques that appeared in all submissions across the entire experiment and rank ordered their “potential” in terms of the highest score attained across multiple solutions using that approach and found that each of the final submissions of those in Open/Disclosure

⁶The significance is higher in unconditional comparisons given the combination of greater individual experimentation and greater participation in Closed/Nondisclosure.

were above the median-scoring approach. The Herfindahl measuring the concentration of solutions across different approaches is also 52% higher, at 0.149 in Open/Disclosure versus 0.0986 in Closed/Nondisclosure. Also consistent with less experimentation and diversity in Open/Disclosure, final solutions appeared in three programming languages in that regime (C#, C++, and Java), whereas in Closed/Nondisclosure 8.7% of final solutions also came from two additional languages (Python and Visual Basic).

D. *Comments on the Mixed Regime*

While the design of our experiment is best geared to evaluating cross-sectional differences between open and closed regimes, we ran a mixed regime to guard against the possibility of eccentric results, given that our design emphasized minimum replication and maximal size of group assignment. The results in the mixed regime appear to corroborate the general patterns found in open and closed regimes. Both the descriptive patterns of submissions and scores, and development of novel combinations of techniques appear to be “between” the patterns observed in open and closed regimes, at least in broadest brush strokes. While the experiment is not designed to evaluate dynamic patterns, several descriptive facts bear noting. For example, while the mixed regime is in between the open and closed regimes in its generation of performance and diversity, the results hardly confirm that the mixed regime is a simple “average” of the open and closed regimes. For example, the mixed regime engenders far more active participation in its second (open) week, in terms of participation and submissions, than did the open regime during this second week. Further, the mixed regime appears to have a higher trajectory of maximal performance in the second week than either the open or closed regime. As a descriptive fact, it also appears that the level of experimentation, as measured by the number of unique combinations assayed also ascends quickest in the second week, among all regimes. Therefore, while the mixed regime serves its purpose of affirming results in open and closed regimes are not driven by eccentric results, at the same time the patterns begin to suggest that performance in an open regime in the

second week may not have been independent of the fact that the knowledge accumulation process during the first week was closed. Thus, the patterns begin to suggest a non-ergodic nature of knowledge accumulation with bearings on subsequent patterns of development.

V. Conclusions

In this paper, we devise an experimental approach to allow us to investigate effects of an open regime in which all intermediate solutions were disclosed and accessible to all subjects working on developing solutions to an algorithmic innovation problem that was amenable to cumulative innovation. Our analysis compared outcomes in this open regime with those in a closed regime in which no solutions were disclosed until the end of the experiment. Subjects in our experiment created solutions to a challenging bioinformatics problem that both industrial and academic labs face, and one that has been subject to a process of cumulative innovation outside of our experiment. Our subjects possessed relevant aptitudes and a mix of skills to address the problem at hand. Thus, while there are unavoidable costs of departing from a natural context of innovation, we attempted to minimize these costs in conveying experimental methods to the evaluation of this policy question shaping scientific and technical progress (Marburger, 2005; Azoulay, 2012).

Our experimental design has the further advantage of overcoming the often unavoidable challenge of studies of naturally occurring contexts in which regimes under comparison are typically not entirely independent, as it is difficult to establish that scientists or other sorts of innovators drawn to different regimes are truly drawn from independent pools of prospective participants or “risk sets.” Crucially, our design allows us to not only maintain independent groups who are then exposed to distinct institutional regimes, but we also are able to precisely match the risk sets on the basis of their abilities and randomize on other characteristics. We were also able to provide novel and precise ways of measuring both performance attained or solution quality and choices of technical approaches in the alternative regimes.

The chief mechanisms shaping technical performance produced by intermediate disclosures related to incentives and “learning” of a sort. The lower appropriability afforded by freer disclosures coincided with drops in participation and development activity, consistent with longstanding theories of economic incentives to make investments in innovation (e.g., Nelson, 1959; Arrow, 1962). (These patterns perhaps deserve special emphasis in the context of this experiment, as such a comparison is typically not possible without explicitly observing matched risk sets of prospective entrants.) Particularly striking is the magnitude of drops in incentives and participation.

We might expect freer, intermediate disclosures to also produce large drops in incentives in contexts beyond just this study. The practical limits of fully recognizing and rewarding upstream contributions should only be made more difficult in a system of intermediate disclosures; intermediate disclosures may relate to a wider range and greater number of incomplete and possibly less standardized and less vetted works. More subtly, attribution may become inherently more difficult where a final work is the result of the contributions of a sequence of accumulating intermediate disclosures, analogous to problems of attribution in teams, for example. Nonetheless, drops in incentives might at least be partially attenuated where institutions might somehow be designed to better protect the appropriability, interests, motivations, rewards and recognition of innovators (in some way that does not impede disclosures, to the extent that is possible).⁷ Further, any drop in incentives might to some degree be attenuated outside a single-problem quality ladder, as studied here, for example participants being able to differentiate their innovations and create complementary

⁷For example while open source development projects are staffed by volunteer and part-time contributors (exerting relative low efforts) they are drawn from a massive pool of prospective participants from a global and substantially larger than what most firms can employ. In addition participation is often driven by intrinsic motivations and an ability to directly use modifications in one’s own work or education. Open source projects also benefit from nuanced informal governance (e.g., culture, trust and norms) to implement reputational payoffs, along with more formal mechanisms such as “signatures” in the code. Similarly large pools of contributors are available for Wikipedia with analogous concerns. By contrast, the Bermuda Principles in the HGP reflected the decision of project organizers and funding agencies (and by extension, public authorities, more generally) to actively subsidize intermediate disclosures and implementing incentives for these intermediate disclosures by both promising continued funding for the projects and threatening sanctions for non-participation.

technologies (Bessen and Maskin, 2009).

Notwithstanding the drop in incentives and participation observed with disclosures of intermediate works in the experiment, the negative incentive effects in this case were outweighed by positive learning effects. “Learning” here is not just the result of knowledge transmissions *per se*; the signals generated by open disclosures coordinated and directed development and inventive activity towards more highly performant technical approaches. Therefore, while subjects tended to conserve their efforts, when they did invest their efforts, participants by and large expected to achieve higher returns from adding to already established approaches than from attempting to pioneer altogether brand new approaches to the problem at hand. This meant not only focusing on existing performant approach, but also choosing to pursue far less experimentation altogether.

This tendency to conform and build on existing successful approaches might be expected in contexts beyond that of this experiment. For example, the result is analogous to the more general notion of the emergence of technological trajectories or scientific paradigms, which create powerful incentives to conform to existing innovation pathways in solving a particular problem (Kuhn, 1962, Dosi 1982). While convergent outcomes were more productive in our particular innovation problem, it is possible to imagine innovation and development contexts in which disclosures may experience more frictions and “learning” of the sort described here could converge on an inferior approach and technical pathway (e.g., David, 1985), i.e. participants in our open regime converged on the “globally” best solution, rather than converging on some “local” inferior approach. Features of our context might have been especially congenial to converging approaches. For example, the problem addressed here was particularly given to cumulative innovation, in the sense it was possible to aggregate multiple optimization techniques. The disclosure environment was also relatively frictionless. Thus, while intermediate disclosures may tend to nudge innovation efforts towards already existing approaches, there may be some ability to moderate this effect. Apart from instituting some frictions (effectively limiting intermediate disclosures), it is also plausible that greater diversity and

experimentation might be “seeded” by adding still more heterogeneity or greater numbers of participants than what was observed in our experiment and by removing the constraint of being on a single quality ladder.⁸

Beyond contributing to the existing literature on the effect of supporting institutions on cumulative innovation, our paper also raises important questions for policy makers responsible for innovation. Given modern scientific and technological progress in a range of domains depends on cumulativeness of knowledge, with current innovators heavily relying on the discoveries of earlier efforts of others, the design of policies that enable both investment in innovation and disclosure to others will be increasingly important for economic growth.

⁸For example in the HGP, entirely all follow on development was in highly differentiated analysis of the data sets that were produced, therefore entirely avoiding the kinds of substitution between upstream and downstream developers.

REFERENCES

- Acemoglu, Daron. "Diversity and Technological Progress," Ch. 6 in, *The Rate and Direction of Inventive Activity Revisited*, J. Lerner and S. Stern, eds (Chicago: University of Chicago Press, 2012).
- Aghion, Philippe, Mathias Dewatripont, and Jeremy Stein, 2008. "Academic freedom, private-sector focus, and the process of innovation," *RAND Journal of Economics*, 39 (2008), 617-635.
- Aghion, Philippe, Mathias Dewatripont, Julian Kolev, Fiona Murray, and Scott Stern. "The Public and Private Sectors in the Process of Innovation: Theory and Evidence from the Mouse Genetics Revolution." *American Economic Review*, 100 (2010), 153–58.
- Aghion, Philippe, and Peter Howitt. "A Model of Growth Through Creative Destruction," *Econometrica*, 60 (1992), 323-351.
- Allen, Robert. "Collective Invention." *Journal of Economic Behavior and Organization*, 4 (1983), 1– 24.
- Altschul, Stephen, Warren Gish, Webb Miller, Eugene Myers, and David Lipman. "Basic local alignment search tool". *Journal of Molecular Biology* 215 (1990), 403–410.
- Arrow, Kenneth. (1962), "Economic Welfare and the Allocation of Resources for Innovation," in *The Rate and Direction of Inventive Activity*, Nelson R., ed., (Princeton: Princeton University Press, 1962).
- Azoulay, Pierre. "Research efficiency: Turn the scientific method on ourselves. Nature," *Nature*, 484 (2012), 31-32.
- Bikard, Michael. "Is Knowledge Trapped Inside the Ivory Tower? Technology Spawning and the Genesis of New Science-Based Inventions" unpublished manuscript, (2012).
- Boudreau, Kevin. "Open platform strategies and innovation: Granting access vs. devolving control," *Management Science*, 56 (2010), 1849–1872.
- Chesbrough, Henry. *Open Innovation*. (Boston: Harvard Business Press, 2006).
- Coase, Ronald. "The Problem of Social Cost," *Journal of Law and Economics*, 3 (1960), 1-44.
- Contreras, Jorge. "Bermuda's Legacy: Policy, Patents, and the Design of the Genome Commons," *Minnesota Journal of Law, Science & Technology*, 12 (2011), 61-125.
- Dasgupta, Partha, and Paul David. "Toward a new economics of science," *Research Policy*, 23 (1994), 487-521.
- David, Paul. "Clio and the Economics of QWERTY," *American Economic Review*, 75 (1985), 332–337.

- Deogirikar, A. Stebbin, M. “Seeking Outstanding 'Open Science' Champions of Change.” White House Office of Science and Technology Policy Press Release, URL: <http://www.whitehouse.gov/blog/2013/05/07/seeking-outstanding-open-science-champions-change> (2013).
- DiNardo, John, and Justin Tobias. “Nonparametric Density and Regression Estimation” *Journal of Economic Perspectives*, , 15 (2001), 11-28.
- Farrell, Joseph. “Cheap talk, coordination, and entry.” *The RAND Journal of Economics*. 18 (1987), 34-39.
- Fauchart, Emmanuelle, and Eric von Hippel. “Norms-Based Intellectual Property Systems: The Case of French Chefs,” *Organization Science*, 19 (2008),187-201.
- Furman, Jeffrey, and Scott Stern. “Climbing atop the Shoulders of Giants: The Impact of Institutions on Cumulative Research,” *American Economic Review*, 101 (2011): 1933–1963.
- Galasso, Alberto, and Mark Schankerman. “Patents and Cumulative Innovation: Causal Evidence from the Courts,” CEP Discussion Paper No. CEPDP1205, (2013).
- Gans, Joshua and Fiona Murray. “Funding Scientific Knowledge: Selection, Disclosure and the Public-Private Portfolio,” (2012) Ch. 1 in, *The Rate and Direction of Inventive Activity Revisited*, J. Lerner and S. Stern, eds (Chicago: University of Chicago Press, 2012).
- Green, Jerry and Suzanne Scotchmer. “On the Division of Profit in Sequential Innovation,” *The RAND Journal of Economics*, 26 (1995), 20-33.
- Gowers, Timothy, and Michael Nielsen. “Massively collaborative mathematics,” *Nature*. 461 (009), 879–881.
- Haeussler, Carolin, Lin Jiang, Jerry Thursby, and Marie Thursby. “Specific and General Information Sharing Among Academic Scientists,” NBER Working Paper Series 15315, (2009).
- Heller, Michael, and Rebecca Eisenberg. “Can patents deter innovation? The anticommons in biomedical research,” *Science*. , 280 (1998), 698-701.
- Huang, Kenneth, and Fiona Murray. "Does patent strategy shape the long-run supply of public knowledge? Evidence from human genetics," *Academy of Management Journal*, 52 (2009), 1193-1221.
- Jaffe, A. B., M. Trajtenberg, and R. Henderson, “Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations,” *The Quarterly Journal of Economics*, 108 (1993), 577–598.

- Justel, Ana, Daniel Peña, and Ruben Zamar. “A multivariate Kolmogorov-Smirnov test of goodness of fit,” *Statistics & Probability Letters*, 35 (1997), 251-259.
- Kitch, Edmund, “The Nature and Function of the Patent System,” *Journal of Law and Economics*, 20 (1977), 265-290.
- Kuhn, Thomas. *The Structure of Scientific Revolutions*. (Chicago: University of Chicago Press, 1962).
- Lacetera, Nicola. “Different Missions and Commitment Power in R&D Organization: Theory and Evidence on Industry-University Relations,” *Organization Science*, 20 (2009), 565-582.
- Lakhani, Karim, Kevin Boudreau, Poh-Ru Loh, Lars Backstrom, Carliss Baldwin, Eric Lonstein, Michael Lydon, Alan MacCormack, Ramy Arnaut, and Eva Guinan. “Prize-based experiments can provide solutions to computational biology problems.” *Nature Biotechnology*, 31 (2013), 108-111.
- Lerner, Joshua and Jean Tirole. “The economics of technology sharing: Open source and beyond,” *The Journal of Economic Perspectives*, 19 (2005), 99-120.
- Lessig, Lawrence. *Remix: Making Art and Commerce Thrive in the Hybrid Economy*. (New York: Penguin Books, 2009).
- Lin, Thomas. 2012. Cracking Open the Scientific Process. *The New York Times*. Jan 16, 2012.
- Mansfield, Edwin. “How rapidly does new industrial technology leak out?” *Journal of Industrial Economics*, 34 (1985), 217–223.
- Marburger, John. “Wanted: Better Benchmarks.” *Science*, 308 (2005), 1087.
- Maas, Han, and Eric-Jan Wagenmakers. “A Psychometric Analysis of Chess Expertise.” *The American Journal of Psychology*, 118 (2005), 29-60.
- Moon, Seongwuk . How does the management of research impact the disclosure of knowledge? Evidence from scientific publications and patenting behavior. *Economics of Innovation and New Technology*, 20 (2011), 1–32.
- Moser, Petra. “How Do Patent Laws Influence Innovation? Evidence from Nineteenth-Century World Fairs,” *The American Economic Review*, 95 (2005), 1214-1236.
- Mukherjee, Arijit, and Scott Stern, “Disclosure or secrecy? The dynamics of Open Science,” *International Journal of Industrial Innovation*, 27 (2009), 459-462,
- Murray, Fiona., Philippe Aghion, Mathias Dewatripont, M., Kolev, J., Stern, S. “Of Mice and Academics: Examining the Effect of Openness on Innovation,” NBER Working Paper Series 14819, (2009).
- Murray, Fiona, and Scott Stern. “Do Formal Intellectual Property Rights Hinder the Free Flow of

- Scientific Knowledge? An Empirical Test of the Anti-Commons Hypothesis.” *Journal of Economic Behavior and Organization*, 63 (2007), 648–687.
- Nelkin, D. “Intellectual property: the control of scientific information,” *Science*, 216 (1982), 704-708.
- Nelson, Richard. "The simple economics of basic research," *Journal of Political Economy*, 67 (1959), 297-306.
- Nelson, Richard. "Uncertainty, Learning, and the Economics of Parallel Research and Development." *Review of Economics and Statistics*, 43 (1961), 351-368.
- Nuvolari, Alessandro. 2004. “Collective Invention during the British Industrial Revolution: The Case of the Cornish Pumping Engine.” *Cambridge Journal of Economics*, 28 (2004), 347– 63.
- Royal Society. “Science as an open enterprise,” The Royal Society Science Policy Centre report 02/12, (2012).
- Romer, Paul. “Endogenous Technological Change,” *Journal of Political Economy*, 98 (1990).
- Rosenberg, Nathan. *Perspectives on Technology* (Cambridge: Cambridge University Press, 1976).
- Rysman, Marc, and Timothy Simcoe. “Patents and the Performance of Voluntary Standard-Setting Organizations,” *Management Science*, 54 (2008), 1920-1934.
- Scotchmer, Suzanne. *Innovation and Incentives*, (Cambridge: The MIT Press, 2004).
- Silverman, Bernard. *Density Estimation for Statistics and Data Analysis*. (London: Chapman & Hall, 1986).
- Simcoe, Timothy. “Explaining the Increase in Intellectual Property Disclosure” in *The Standards Edge: The Golden Mean* (Bolin Group, 2007).
- Stephan, Paula. “The economics of science,” *Journal of Economic Literature*, 34 (1996), 1199-1235.
- Stern, Scott. *Biological Resource Centers: Knowledge Hubs for the Life Sciences*, (Washington D.C.: The Brookings Institution Press, 2004).
- Weitzman, Martin. “Recombinant growth,” *The Quarterly Journal of Economics*, 113 (1998), 331–360.
- Williams, Heidi. “Intellectual Property Rights and Innovation: Evidence from the Human Genome,” *Journal of Political Economy*, 121 (2013), 1-27.

TABLES

TABLE I
OLS REGRESSIONS OF PARTICIPATION AND ACTIVITY

Dependent Variable:	<i>Participated_i</i>			<i>NumSubmissions_i</i>			<i>NumSubmissions_i, Conditional on Participating</i>		
	Model:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
<i>OpenDisclosure_i</i>	-.05*	-.06*	-.06**	-.90***	-.95***	-.95***	-3.9***	-4.0***	-3.8***
	(.03)	(.04)	(.03)	(.32)	(.26)	(.26)	(1.1)	(1.0)	(1.2)
<i>SkillRating_i</i>		.06***	.07**		.34**	.56***		.16	.37
		(.02)	(.03)		(.13)	(.21)		(.45)	(.77)
<i>OpenDisclosure_i × SkillRating_i</i>			-.01			-.41*			-.49
			(.04)			(.24)			(.92)
Constant	.19***	.19***	.19***	1.3***	1.3***	1.4***	6.9***	6.9***	6.8***
	(.02)	(.02)	(.02)	(.29)	(.25)	(.25)	(.92)	(.90)	(1.04)
Adj R-Squared	.00	.03	.03	.02	.03	.04	.12	.10	.09

Columns (1) through (6) report models that are estimated using data from the 489 subjects in Open/Disclosure and Closed/Nondisclosure. Subjects in each regime are matched on *SkillRating* and otherwise randomized. Columns (7) through (9) reports estimates based on the 79 actively participating individuals in those regimes. Estimates are from OLS. Numbers in brackets are standard errors, estimated by bootstrapping. The annotations of *, **, and *** indicate statistical significance at the 10%, 5% and 1% levels, respectively.

TABLE II
COMPARISON OF MEANS AND STANDARD DEVIATIONS OF ABILITY MEASURES

	Mean (μ)					Standard Deviation (σ)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Pop. Mean	Closed Active Participants	Open Active Participants	$\mu_{open} -$ μ_{closed}	Std. Error	Pop. Std. Dev.	Closed Active Participants	Open Active Participants	$\sigma_{open}^2 /$ σ_{closed}^2
<i>SkillRating</i>	.00	.26	.48	.23	(.24)	1.0	1.1	1.1	1.06
<i>NumPastParticipations</i>	27.3	41.3	43.2	.1	(13)	43.0	8.0	9.9	1.1
<i>YearJoinedPlatform</i>	2006.5	2006.0	2006.0	0.1	(.39)	1.9	1.7	1.6	0.9

Columns (1) and (2) report mean and standard deviation for the entire experimental population of 733 subjects (equivalent to those individual groups, given the assignment procedure). Columns (2) and (7) report values among those subjects choosing to participate within Closed/Nondisclosure. Columns (3) and (8) report values of participants in Open/Nondisclosure. Column (4) and (5) reports the differences and standard error of this difference in means within brackets. These differences are not statistically significant at customary levels. Column (9) reports the ratio of estimated variance in the two regimes. This ratio is not statistically significant at customary levels.

TABLE III
COMPARISON OF PROPORTIONS OF PARTICIPANT CHARACTERISTICS

	(1)	(2)	(3)	(4)	(5)
	Population Proportion	Closed Active Participants	Open Active Participants	$\mu_{open} - \mu_{closed}$	Std. Error
<u>Technical Area of Primary Interest</u>					
<i>Software</i>	34%	52%	39%	-13%	(11%)
<i>Games</i>	4%	2%	6%	4%	(4%)
<i>Web</i>	2%	2%	3%	1%	(4%)
<i>Other</i>	60%	43%	52%	8%	(11%)
<u>Country</u>					
<i>India</i>	20%	7%	12%	5%	(6%)
<i>USA</i>	16%	15%	15%	0%	(8%)
<i>Russia</i>	9%	15%	9%	-6%	(8%)
<i>China</i>	9%	11%	12%	1%	(7%)
<i>Other</i>	46%	52%	52%	-1%	(12%)

Column (1) reports the probability of a given individual in the entire experimental population of 733 subjects (equivalent to those individual groups, given the assignment procedure), equivalent to the proportion in the population. Columns (2) and (3) report values among those subjects choosing to participate within Closed/Nondisclosure and Open/Nondisclosure. Column (4) reports the difference in means. Column (5) reports the standard error of this difference in brackets. All differences are not statistically significant at customary levels.

FIGURES

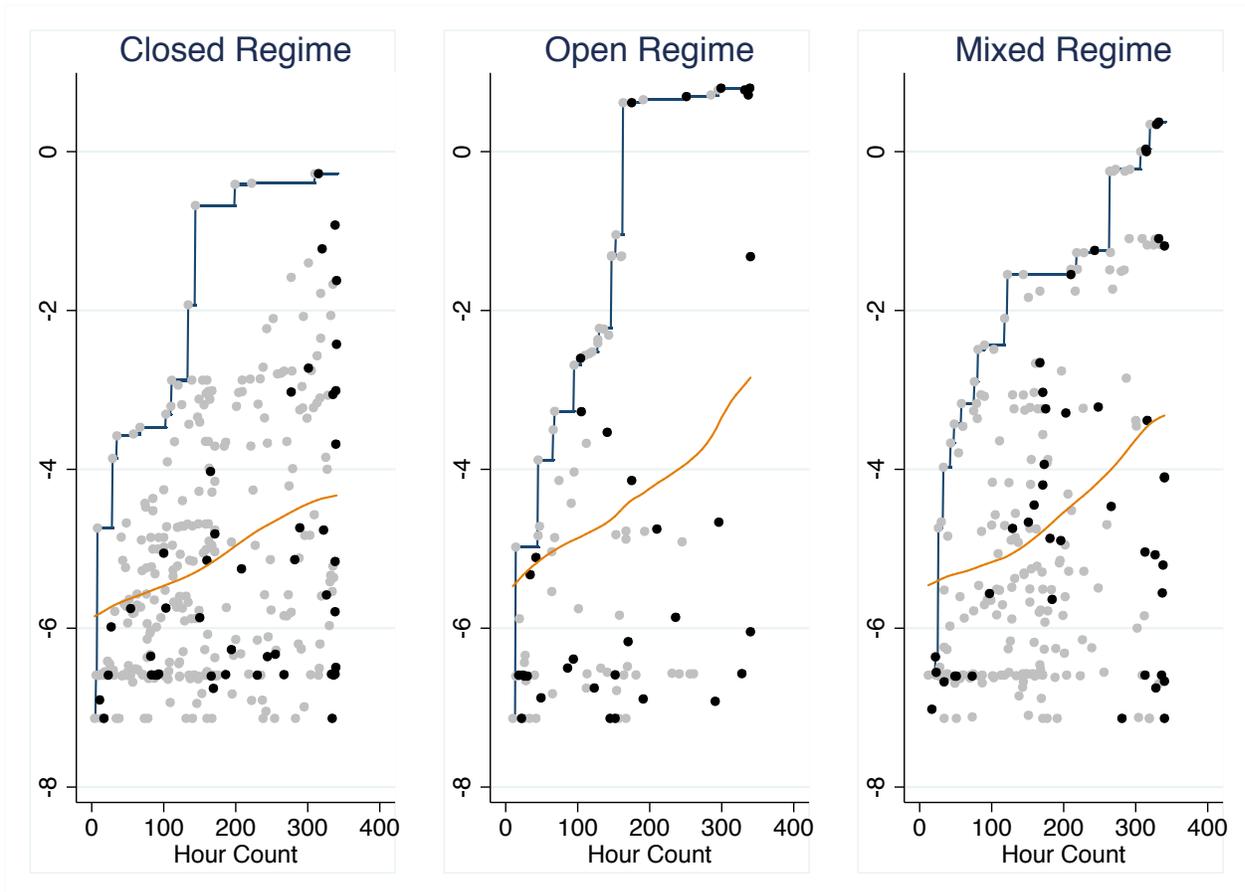


FIGURE I

The Cumulative Innovation Process: Submissions and Solution, By Regime

The figure plots the incidence of intermediate solution submissions is indicated by grey dots and final submissions by individuals indicated by black dots. The horizontal axis indicates the number of hours passed over the course of the 2-week experiment. The vertical axis records the quantitative assessment of the quality (an amalgam of speed and accuracy) assigned to each submission on the basis of an automated test suit. The black line is the maximal frontier of scores attained over time in each regime is traced in each chart. The red line is a moving mean.

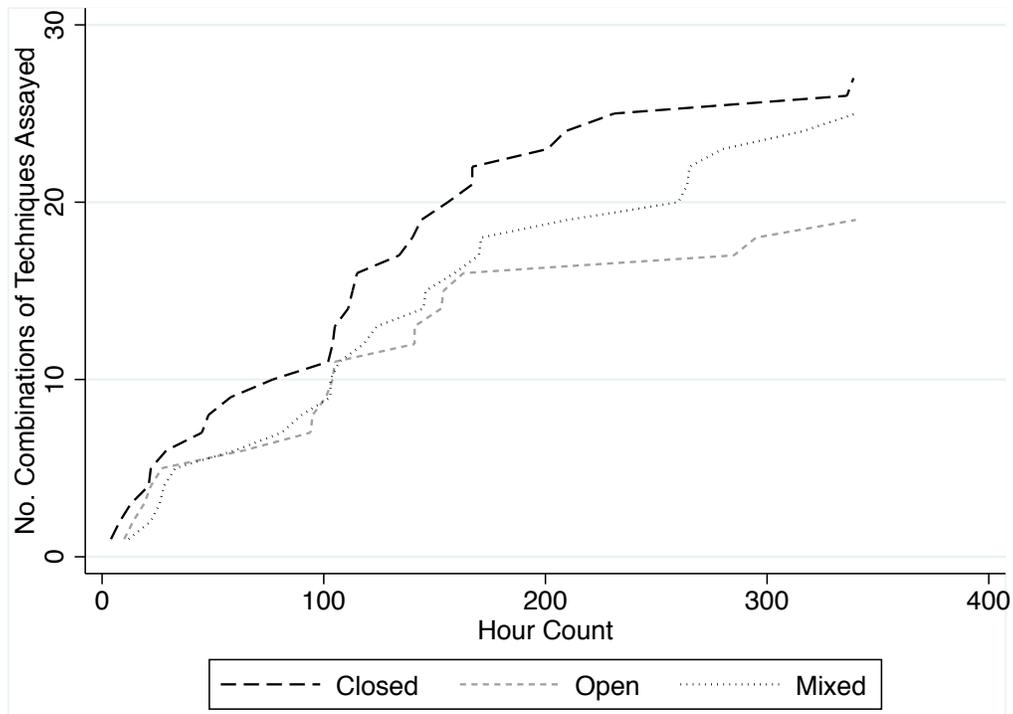


FIGURE II

The Extent of Experimentation: Cumulative Count of Unique Approaches, by Regime.

Each of 654 solutions was coded according to their use of ten elemental optimization techniques. The figure plots the accumulation of first instances in which a given combination of approaches was used within a given regime. The vertical axis indicates the total count of unique approaches. The horizontal axis indicates the number of hours passed over the course of the 2-week experiment.

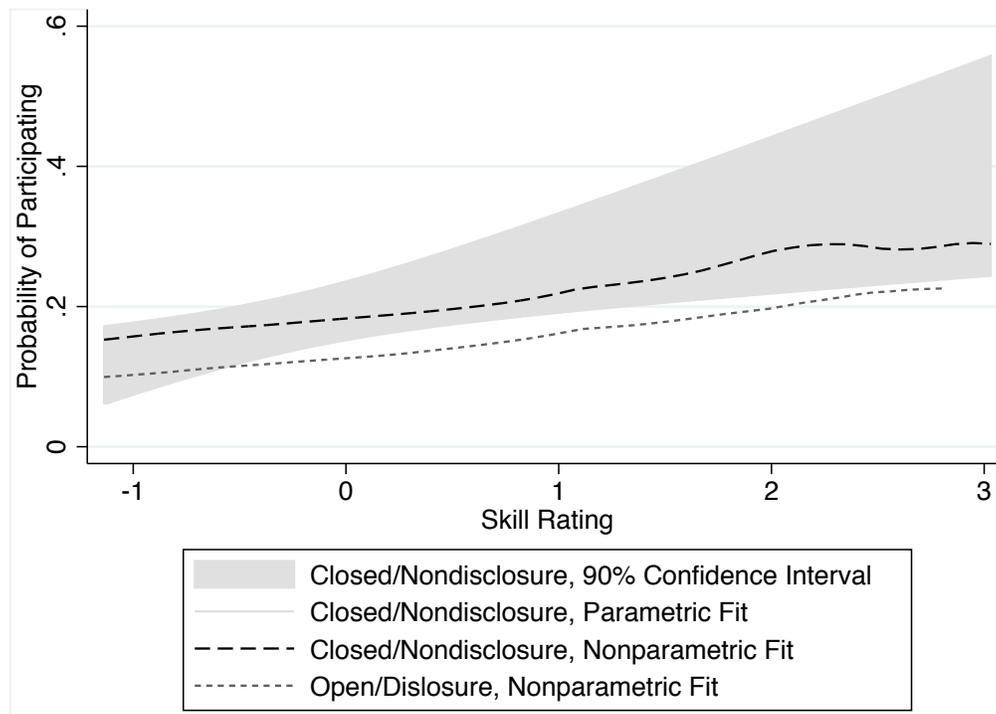


FIGURE III
Willingness to Participate: Probability of Participating, by Regime, by Skill

The figure plots a regression line of probabilities of individuals choosing to participate (i.e., making at least one solution submission) in Open/Disclosure and Closed/Nondisclosure, stratified by skill level. The grey area shows the 90% confidence interval estimated from a linear OLS model for the Closed/Nondisclosure regime. The black dashed line shows a flexible, non-parametric estimator based a locally weighted quadratic polynomial fit. The grey short dashed line is the analogous non-parametric estimator of the relationship in the Open/Disclosure regime.

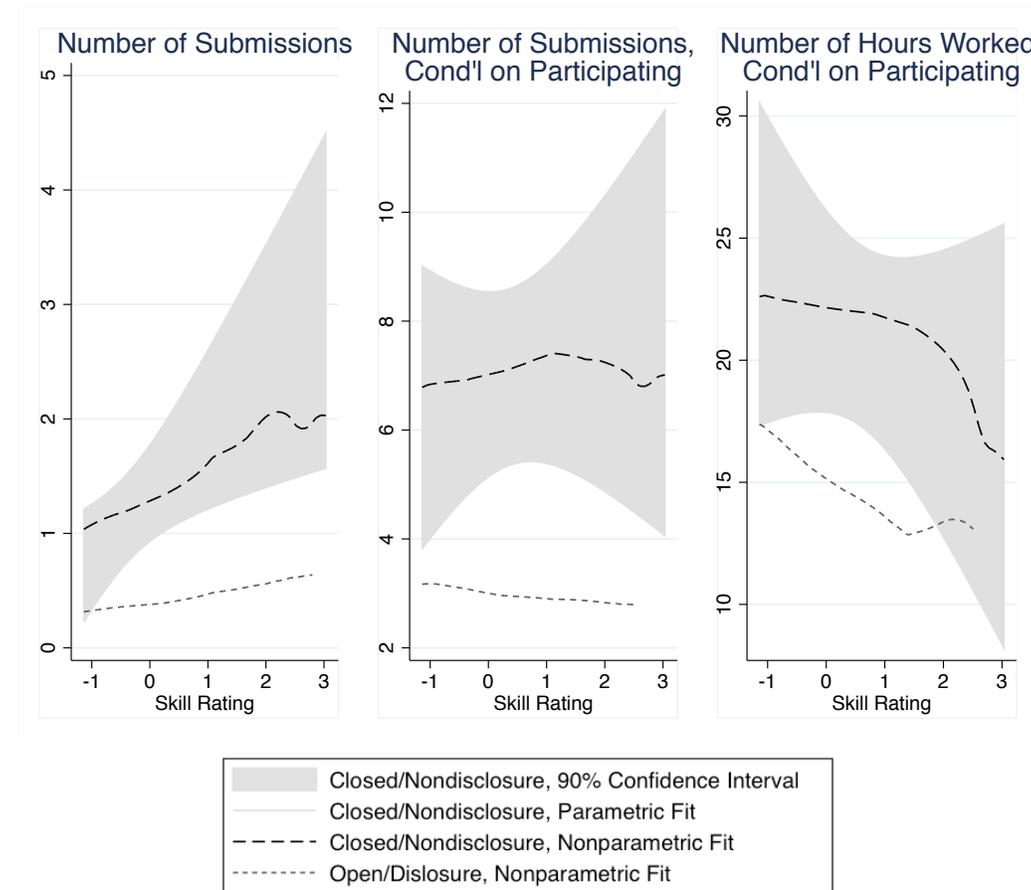


FIGURE IV

Willingness to Exert Effort: Measures of Activity and Effort, by Regime, by Skill, Unconditional and Conditional on Participating

Each panel of the figure plots a regression line of measures of effort and activity in Open/Dislosure and Closed/Nondisclosure, stratified by skill level. The first panel plots the number of submissions of solutions (an indication of trial-and-error activity) for all subjects, including non-participants. The second panel plots the number of submissions of solutions, but only for participants (i.e., those who submitted at least once). The third panel plots number of hours worked, as self-reported by participants. The grey area shows the 90% confidence interval estimated from a linear OLS model for the Closed/Nondisclosure regime. The black dashed line shows a flexible, non-parametric estimator based a locally weighted quadratic polynomial fit. The grey short dashed line is the analogous non-parametric estimator of the relationship in the Open/Dislosure regime.

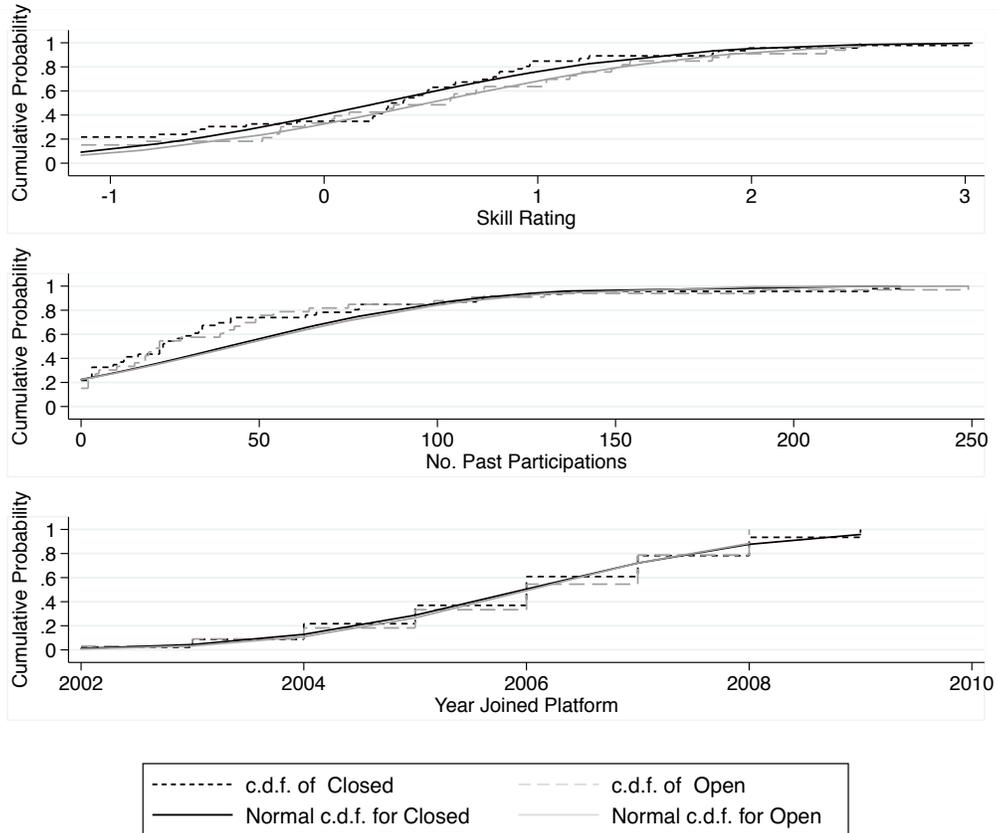


FIGURE V

Nearly Equivalent Ability Distributions Among Participants: Distribution of Ability Among Active Participants, by Regime

The three panels plot measures of the skills distribution of the subset of subjects who chose to actively participate. Each panel records both empirical cumulative distributions, along with fitted parametric normal approximations, for both Open/Disclosure and Closed/Nondisclosure. The first panel plots the Elo-based measure of ability in solving algorithmic problems, *SkillRating*. The second panel plots the number of instances in which a participants participated in analogous events on the platform in the past, prior to the experiment, *NumPastParticipations*. The third panel plots the cohort of participants, *YearJoinedPlatform*.

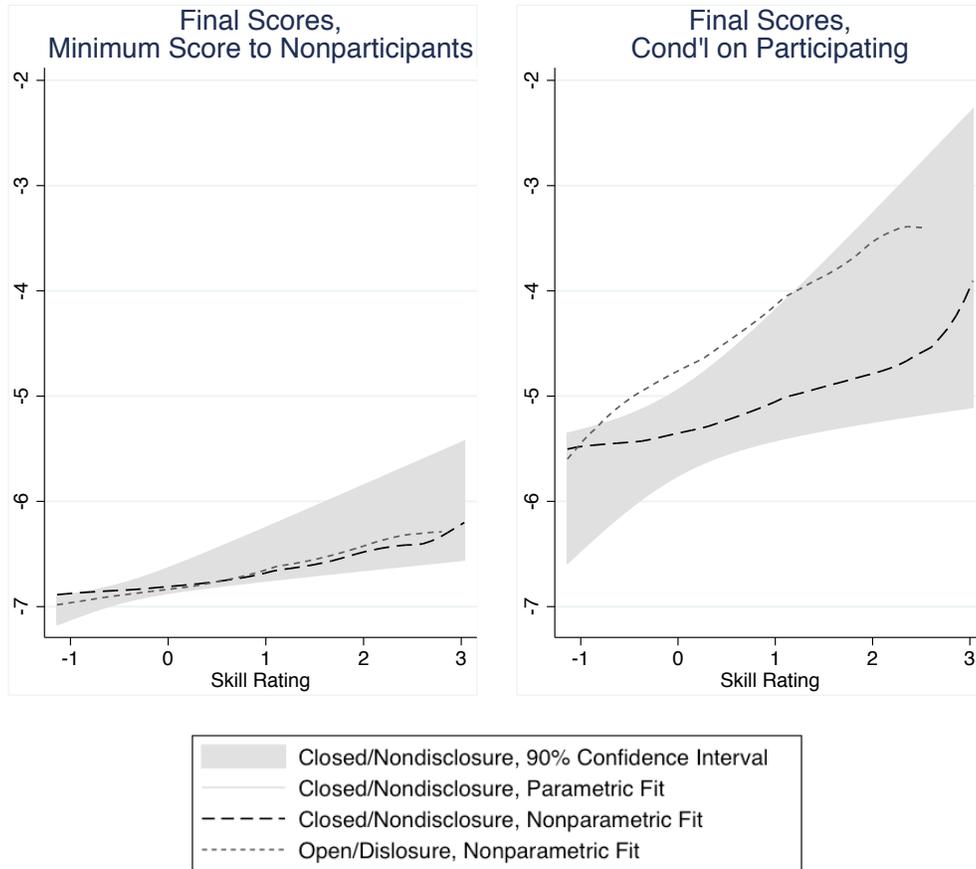


FIGURE VI

Performance: Problem-Solving Score, by Regime, by Skill, Unconditional and Conditional on Participating

Each panel of the figure plots a regression line of performance, as measured by final scores, in Open/Disclosure and Closed/Nondisclosure, stratified by skill level. The grey area shows the 90% confidence interval estimated from a linear OLS model for the Closed/Nondisclosure regime. The black dashed line shows a flexible, non-parametric estimator based a locally weighted quadratic polynomial fit. The grey short dashed line is the analogous non-parametric estimator of the relationship in the Open/Disclosure regime.

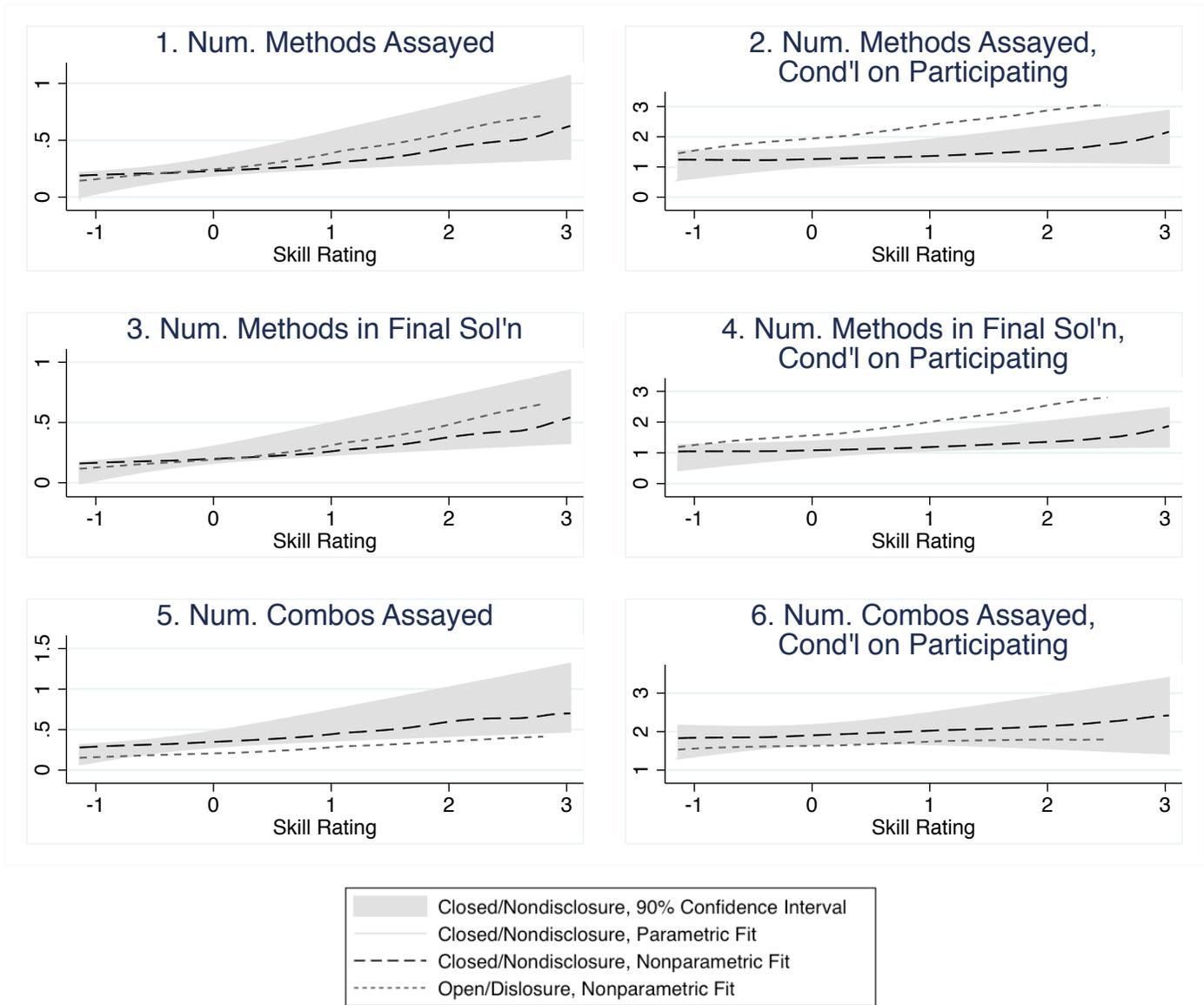


FIGURE VII

Technical Approaches and Experimentation: Number of Techniques Assayed and Implemented in Final and Number of Combinations Tried, by Regime, by Skill, Unconditional and Conditional on Participating

The first and second panels plot numbers of techniques tried across all submissions by a subject, unconditional and conditional on participating (i.e., submitting at least once). The third and fourth panels plot numbers of techniques implemented within the final solution, unconditional and conditional on participating. The fifth and sixth panels plot the number of new combinations of techniques tried by subjects across all their submissions, unconditional and conditional on participating. The grey area shows the 90% confidence interval estimated from a linear OLS model for the Closed/Nondisclosure regime. The black dashed line shows a flexible, non-parametric estimator based on a locally weighted quadratic polynomial fit. The grey short dashed line is the analogous non-parametric estimator of the relationship in the Open/Disclosure regime.