Research Policy 44 (2015) 4–19

Contents lists available at ScienceDirect

Research Policy

journal homepage: www.elsevier.com/locate/respol

“Open” disclosure of innovations, incentives and follow-on reuse: Theory on processes of cumulative innovation and a field experiment in computational biology

Kevin J. Boudreau[a,*], Karim R. Lakhani[b]

[a] London Business School, Strategy & Entrepreneurship Department, Regent's Park, London NW1 4SA, UK
[b] Harvard Business School, Technology and Operations Management & NASA Tournament Lab at Harvard Institute for Quantitative Social Science, MA, USA

ARTICLE INFO

Article history:
Received 5 November 2013
Received in revised form 1 August 2014
Accepted 11 August 2014
Available online 5 October 2014

Keywords:
Innovation
Disclosures
Search
Incentives
Policy
Open innovation

ABSTRACT

Most of society's innovation systems – academic science, the patent system, open source, etc. – are “open” in the sense that they are designed to facilitate knowledge disclosure among innovators. An essential *difference* across innovation systems is whether disclosure is of intermediate progress and solutions or of completed innovations. We theorize and present experimental evidence linking intermediate versus final disclosure to an ‘incentives-versus-reuse’ tradeoff and to a transformation of the innovation search process. We find intermediate disclosure has the advantage of efficiently steering development towards improving existing solution approaches, but also has the effect of limiting experimentation and narrowing technological search. We discuss the comparative advantages of intermediate versus final disclosure policies in fostering innovation.

1. Introduction

How do disclosure policies governing the reuse of knowledge, technology and innovations, once developed, affect the rate and direction of inventive activity? Consider, for example, the implementation of the “Bermuda Principles” by the Human Genome Project (HGP). In return for stable, guaranteed funding, more than one thousand research scientists representing more than 30 research laboratories in at least 19 countries agreed to following a procedure of releasing sequence data into the public domain within 24 hours of discovery (Contreras, 2011). Conducted over a thirteen-year period, the HGP was one of the most ambitious, large-scale, scientific efforts in modern times. This process of near instantaneous disclosure of findings and methods, intended to enable investigators to build on each others' results and coordinate in order to more rapidly advance the frontiers of scientific knowledge, yielded the structure of the human genome. This immediate disclosure and public sharing of scientific results (sequences) was a significant departure from the usual academic practice of releasing results and analyses in the form of a published scientific journal article. (These intermediate outputs would later be used in more typical scientific publications down the road). It was also in stark contrast to the patenting and contracting strategy pursued by Celera, the for-profit firm that was racing against the HGP consortium.

As the example illustrates, there is a range of institutions in which innovation can be governed (see Table 1 for examples). This includes, for example, the patent system, academic science, open source and creative commons licenses, and also ad hoc contracted frameworks, as in the HGP. Each framework or innovation system treats knowledge and technology disclosure (e.g., sharing, spillovers, transfers and reuse more generally) according to its own particular rules and procedures (Table 1). Rather than focus on the many details of disclosure policies within a given innovation system, this paper considers an essential difference across innovation systems: *final* versus *intermediate* disclosure policies. Here, we investigate how final and intermediate disclosure policies – this key difference distinguishing innovation systems – shapes innovation. Thus, we attempt to better understand the comparative advantages of these approaches and the innovation systems that embody them.

Traditional institutions tend to favor final disclosure of an innovation or problem-solving output that is completed or “working,” i.e. after the innovation-related problem solving process is completed, such as patented inventions, working instantiations of designs in product components or machinery used in larger downstream systems, vetted academic contributions in the form

* Corresponding author. Tel.: +44 020 7000 8455; fax: +44 020 7000 8701.
E-mail addresses: kboudreau@london.edu (K.J. Boudreau), k@hbs.edu (K.R. Lakhani).

http://dx.doi.org/10.1016/j.respol.2014.08.001
0048-7333/

Table 1
Illustrative examples of cumulative innovation systems, by disclosure policy.

Governance framework	Unit of innovation output/disclosure	Implementation of incentives	Implementation of disclosure and reuse
Final disclosure			
Patent System	A working invention	Patents confer finite rights of exclusion, which in principle may enable property rights and trade or, alternatively, monopoly supply. (e.g., Arrow, 1962; Nordhaus, 1972; Green and Scotchmer, 1995; Bessen and Maskin, 2009)	Patent award leads to disclosure in the public domain with general public access once patent expires. Prior to this, licensing or imperfect protections enable reuse. (e.g., Kitch, 1977; Green and Scotchmer, 1995; Cohen et al., 2000; Arora et al., 2004; Chesbrough, 2003; Williams, 2013)
Academic Science	A complete research publication	Quality, number of publications and number of citations are the basis of rewards in Science (promotion, status, funding, peer esteem honors and awards). Intrinsic and other sources of motivation can also play a large role. (e.g., Cohen et al., 2000; Stern, 1994; Stephan, 1996; Aghion et al., 2008)	Publications are disclosed in the public domain (i.e., academic journals, working paper databases) whereupon others, conditional on citation, can reuse their content and ideas. (e.g., Cohen et al., 2000; Stephan, 1996; Salter and Martin, 2001; Murray et al., 2009; Bikard, 2014)
Ansari X-Prize for Suborbital Space Flight (and other public contests)	A complete, working solution to the challenge	Cash payoffs and public acclaim are on the basis of rank-ordered outcomes based on preannounced criteria. (e.g., Taylor, 1995; Moldovanu and Sela, 2001; Boudreau et al., 2011; Murray et al. 2012)	Disclosure in prizes may, in principle be dealt with any number of ways. In this case, ownership over the winning technology was retained by the winning solution provider. (e.g., Moser and Nicholas, 2013)
Apple AppStore (and other multi-sided platforms)	An "App"	The platform creates two-sided market-based incentives to make sales to the large number of users of Apple devices. (e.g., Bresnahan and Greenstein, 1999; Rochet and Tirole, 2003; Parker and Van Alstyne, 2010; Boudreau, 2012)	The upstream platform is, by design, a technology intended to be reused and built upon. (e.g., Baldwin and Clark, 2000; Boudreau, 2010; Parker and Van Alstyne, 2010)
Intermediate disclosure			
Open Source Projects	A code contribution, edit or bug report	Projects enlist contributors with intrinsic, own-use or pecuniary motives, in addition to those wishing to learn, affiliate or signal their mastery through high quality contributions. (e.g., Lerner and Tirole, 2002; Lakhani and von Hippel, 2003; Lakhani and Wolf, 2005; Roberts et al., 2006)	General Public licenses (GPL) impose stringent requirements on those distributing GPL-based code to mandatorily disclose and grant rights of access to others. (e.g., Lerner and Tirole, 2002; Lessig, 1999)
Human Genome Project (Bermuda Principles)	The gene sequence.	Public scientific institutions partaking in the HGP were required to accede to Bermuda principles as a condition of research funding. (e.g., Contreras, 2011)	The Bermuda Principles required public disclosure within 24 h of discovery of sequence information. (e.g., Williams, 2013)
Homebrew Computer Club (and other collective invention, "user" and innovator communities)	The idea or technique	Those drawn to participate in informal associations and communities of innovators may be motivated by wide range of reasons, including those related to learning, intrinsic motivation, socialization and many more. "Community" can also initiate an incentive to reciprocate. (e.g., von Hippel, 1988; von Hippel and von Krogh, 2003; Osterloh and Rota, 2007)	The typically informal nature of communities leads to few formal restrictions to disclosure and reuse. However, informal rules and norms and threat of sanction may be the basis for imposing conditions, such as acknowledgments. (e.g., Allen, 1983; Nuvolari, 2004; Fauchart and von Hippel, 2008)
Foldit Protein Folding Platform (and other aggregative, collaborative platforms)	The "fold" or "contribution"	The Foldit interface is devised to entice an intrinsic puzzle-solving response of contributors through an online game. More generally, collaborative and aggregative platforms may mobilize effort through any number of incentives in collaborations. (Khatib et al., 2011; Zhang and Zhu, 2011; Franzoni and Sauermann, 2014)	The assembled database is available to scientists who study highest scoring solutions and who can then proceed to use these intermediate outputs as a basis for developing academic publications.

of published journal articles, artistic or compositional products in some integral form and so on (Table 1). Intermediate and final disclosure are distinguished in the first instance by *timing*. Whereas final disclosures necessarily occur upon (and often considerably *after*) the completion of work, intermediate disclosure occurs continuously. Disclosure is further distinguished by *form*. Final disclosures, by their definition, typically involve some standardized, integral, working and wholly resolved form of solution. By contrast, intermediate disclosure can accommodate a greater breadth or smaller quanta of knowledge, as in partial and negative results, methods, data, progress updates and so forth.

We argue and test two main points. First, we argue that more readily promoting knowledge reuse through intermediate disclosure comes with the cost of diminished incentives, depressing effort and participation – an "incentives-versus-reuse" trade-off. We clarify that this tradeoff is rooted in the timing, form and contractibility of technology and knowledge reuse. Second, intermediate and final disclosure policies produce a qualitative transformation in patterns and scope of "search" across different approaches to addressing an innovation problem – both in terms of choices by individual innovators and overall patterns in the population of innovators. Final disclosure promotes greater independent, "parallel" or uncorrelated experimentation across different innovators (Nelson, 1961; Abernathy and Rosenbloom, 1969; Boudreau et al., 2011); intermediate disclosure produces more coordinated – and possibly convergent and overlapping – choices of solution approaches. These differences are shaped by the prospect of reducing cost and uncertainty in experimentation and by greater signaling (regarding the knowledge frontier and the actions of others) under intermediate disclosure.

Whereas, within the economy, intermediate and final disclosure are associated with entirely different innovation systems, our empirical investigation makes inferences by varying the disclosure policy while holding other features of institutional design constant. To do so, we implemented a field experiment in a controlled, "synthetic" institutional environment. We implemented the experiment on an online platform that was built and customized to incorporate key research design features by TopCoder (a leading developer of contest-based custom software and algorithm solutions).

The design involved comparing randomly-assigned independent groups of individuals working to develop and optimize a bioinformatics algorithm under either disclosure regime. In all, 733 mathematicians, software developers, scientists and data scientists participated over the two-week problem-solving period. Under intermediate disclosure, intermediate solutions developed in the regular trial-and-error development process were instantaneously catalogued and made available for inspection and reuse by other participants within the group. Under final disclosure, intermediate solutions were not disclosed until the end of the two-week development period. Payoffs and rewards were on the basis of rank order of solution performance within each independent group. We observe fine-grained measures of incentives and effort, solution approaches and the technical performance of solutions.

It should be emphasized that the experimental design is intended to reflect a *population* of prospective innovators and how they respond to the prospect of working under a given institutional framework. This creates the need for unusually large comparison groups to be constructed from the 733 participants. A direct consequence and cost of this requirement is minimal replication in the design. We discuss this point at greater length herein. A second point deserving special emphasis is the research is intended to generalize insights in relation to a wide range of innovation systems on the basis of the one (synthetic) institutional context presented here. Therefore, we caution that while the intermediate disclosure regime produces higher quality problem-solving at lower cost in this particular context, this is not a generalizable finding. The key generalizable insights reside in the tradeoffs we document.

The paper is organized as follows. In Section 2, we establish key terms of reference with respect to disclosure policies in innovation systems. In Section 3, we review related literature and develop predictions. The experimental set-up, methods and data collection are described in Section 4. Section 5 describes the data. Section 6 presents results and analysis. Section 7 concludes.

2. The role of disclosure policies in innovation

In this section, we establish our key terms of reference, first discussing the cumulative innovation process and then defining questions of disclosure more precisely.

2.1. "Upstream" knowledge and "downstream" reuse

Scholarship in a range of disciplines has conceived of innovation as a cumulative process whereby the frontiers of knowledge and production possibilities are advanced by successfully solving a series of problems (Kuhn, 1962; Sahal, 1985). In large part, new knowledge, innovation and technical advances are products of a recombinatorial process (e.g. Weitzman, 1998; Fleming, 2001), where existing "upstream" knowledge is built upon and recombined within an on-going stream of cumulative innovations, including those that improve upon the original application and perhaps others that open up new uses (Basalla, 1988). In academic science, for example, this takes the form of new advances building upon and citing existing publications, presentations and exchanges in meetings and seminars, etc. (Dasgupta and David, 1994). In industrial innovation, for example, competitors may learn from and draw on the knowledge and technology of other firms through some combination of licensing, involuntary knowledge spillovers, movement of employees and so forth (e.g., Marx et al., 2009; Rosenkopf and Almeida, 2003). Similarly, various instantiations of open innovation systems rely on participants being able, in their attempts to solve current problems and create new inventions, to reuse prior contributors' knowledge and technology (e.g., Chesbrough, 2003; von Hippel, 2005).

2.2. Disclosure, granting access and devolving control

Where innovating individuals or organizations possess distinct comparative advantages, it will sometimes be productive to involve multiple parties to innovate within the chain of cumulative innovation (Green and Scotchmer, 1995). Therefore, central to cumulative innovation is a need for upstream knowledge and technology to be disclosed in order for downstream innovators to reuse and build upon this work.

Our use of "disclosure" here should be understood as shorthand for the idea of implementing a broader framework[1] in which upstream knowledge and technology are not just disclosed but follow-on innovators are also *granted access* rights (Murray and O'Mahony, 2007; Boudreau, 2010). For example, patents disclose the designs of inventions in the public domain. But patents also confer rights of exclusion to the patent owner; it is through licensing that access is granted for reuse by downstream innovators. Imperfect defensibility of patents can also lead to *de facto* access through "leakage" and involuntary spillovers of knowledge. Analogously, "user innovation" requires that originating technologies and ideas not only be disclosed, but also that users have rights of access and reuse, typically via the "first-sale doctrine" (Katz, 2014), which grants inventors rights to adapt, change and modify existing products without legal encumbrances. Furthermore, beyond providing access via a contractual framework, in the case of physical materials there can be a need for investments in facilities and infrastructure to enable transfer and downstream reuse (Stern, 2004; Furman and Stern, 2011).

Disclosures and access typically impose certain conditions and stipulations, include those related to use, sharing, further development, modification and commercialization (e.g., Gans and Murray, 2012). Stipulations may concern issues such as payments, attribution, responsibilities and restrictions of various kinds. Wholly devolving control rights over upstream knowledge or technology may eliminate such restrictions and conditionality (Boudreau, 2010). Complete devolution of control rights, however, is rare. Even in the case of open source software, rights of access and reuse are only established by acceding to the General Public License or like agreements, which place a great number of conditions on sharing, further development, modification and commercialization (Raymond, 1999).

2.3. Intermediate versus final disclosure policies

Following earlier premises, we contend that all innovation systems are "open", if by "open" we mean they enable disclosure of upstream knowledge and technology for reuse. Although endless

[1] We focus here on disclosure *policies* rather than strategic, voluntary disclosures by individual actors (e.g., Haeussler et al., 2009; Henkel and Baldwin, 2010)

details distinguish different innovation systems (see Table 1), a categorical distinction can be made between systems implementing *final* and *intermediate disclosure* policies. As discussed in the Introduction, innovation systems implementing intermediate or final disclosure can be distinguished by the *timing* and *form* of the disclosure of the innovative output. Final disclosure occurs upon the completion of work (and often considerably after) and typically involves some standardized and integral form of output. By contrast, intermediate disclosure occurs more continuously throughout the innovative problem solving process and can accommodate a greater range and varying quanta of knowledge.[2]

Perhaps the most common and influential example of intermediate disclosure is open source software, in which code, bug reports and test suites are instantly made available for developers to build upon and reuse in successive submissions to the code base (O'Mahony, 2003; von Krogh et al., 2003; von Hippel and von Krogh, 2003). Intermediate disclosure practices have also been implemented in computer hardware (Osterloh and Rota, 2007), Wikipedia (Zhang and Zhu, 2011), synthetic biology (Torrance, 2010), the Polymath Project for creating mathematical proofs (Gowers and Nielsen, 2009) and Netflix's $1MM prize contest to improve its movie rating prediction algorithm, in which intermediate solutions were disclosed in the course of the contest. Intermediate disclosure is not just a modern, Internet-driven phenomenon. Ample case examples from the industrial revolution and early 20th century of particular technological advances describe instances of "collective invention" in which inventors making intermediate disclosures to one another of ideas and techniques propelled advances in blast furnace technology (Allen, 1983), Cornish pumping engines, Bessemer steel, large-scale silk production (Nuvolari, 2004) and aviation technology (Meyer, 2013).

3. Disclosure policies and innovation outcomes

In this section, we consider how implementing final or intermediate disclosure policies may affect innovation. Our discussion is in two parts. We first describe what we refer to as the incentives-versus-reuse tradeoff. We then describe how intermediate versus final disclosure policies qualitatively transform the search process. We develop predictions for the empirical analysis to follow.

3.1. Disclosure policies, incentives and follow-on reuse

Creating *ex ante* incentives to make costly and risky investments and effort while simultaneously encouraging *ex post* or follow-on knowledge reuse (through disclosure) are two goals[3] of any innovation system (see Scotchmer (2004) and Table 1). In any one innovations system, the goals of incentives and reuse can sometimes conflict and other times be complementary. For example, knowledge spillovers and leakage among competing firms facilitates reuse, but may harm incentives (Scotchmer, 2004). In academic science, greater reuse through citations might, by contrast, stimulate efforts and incentives (Stephan, 1996, 2012). In this section, rather than focus on incentives and reuse *within* a given system, we instead consider how differences in disclosure policies *across* systems bear on innovation outcomes.

3.1.1. Intermediate disclosure and contractibility

As is well known, contractibility conditions are tenuous around the transfer and reuse of knowledge and ideas (e.g., Arrow, 1971). It is challenging, under the best of circumstances to assure contractibility in the sense of assuring that disclosure and access conditions are honored and proper rewards are conferred to the originating innovator. The challenge of contractibility is so central to innovation system design that the nuanced rules, laws, procedures and even cultures and customs of any one system can be readily interpreted in this light. For example, multi-party exchange frameworks – as in patent pools, academic publishing, standards organizations and biological research centers – avoid the need for *ad hoc* bilateral negotiations and governance. Informal governance can also supplement formal governance in order to enforce rules of reuse and sharing (Fauchart and von Hippel, 2008). Proprietary platform technologies are even designed from the ground up in a way that enables access and reuse by large numbers of downstream innovators without the need to relinquish control or transfer knowledge of the inner workings of the platform (Boudreau and Hagiu, 2009; Boudreau, 2010; Parker and Van Alstyne, 2010).

Most crucial to our arguments here are the cornerstones of contractability, observability and verifiability (Hart, 1995). Intermediate disclosure degrades observability and contractibility of knowledge disclosure and reuse. This is because the delicate balance in the design of any innovation system can best function to the extent that upstream knowledge being disclosed, transferred and reused takes some completed, working, wholly integral form – simply because this makes the quantum of knowledge transfer more measurable and standardized.

Consider the academic science context, where the research article is the commonly accepted and readily observed "unit of innovative output." Production of research articles is governed by commonly understood requirements for format, content and completeness. These outputs are even certified by the institutions through peer review. Peer review is itself somewhat standardized through a set of regular routines, given standard expectations around the content of research articles. Simple counts of research articles also even act as meaningful measures of journal quality (impact factor as average numbers of citations) and researcher quality (on the basis of publication counts, by journal type). Intermediate disclosure will reduce these levels of observability and verifiability if they imply less standardized, less measurable, less certified, more commingled, more varied forms and proper attribution is made more difficult.

3.1.2. The "incentives-versus-reuse" tradeoff

In considering how incentives and reuse are affected by disclosure policies, prior research has tended to consider these issues separately, rather than at once; and this research also tends to consider these issues within a given innovation system, rather than in relation to different approaches to disclosure across altogether different systems. Here, we describe an "incentives-versus-reuse" tradeoff associated with intermediate and final disclosure, when considering broader comparisons.

With respect to incentives, intermediate disclosure reduces contractibility, as earlier described, reducing upstream innovators' ability to impose conditions and stipulations on reuse. This includes stipulations assuring recognition and rewards to the upstream innovator. Intermediate disclosure therefore reduces incentives, all else being equal.[4] This claim is consistent with the popular

[2] Our consideration of timing here differs from past research on questions of timing and breadth of disclosure (Mazzoleni and Nelson, 1998; Mukherjee and Stern, 2009; Lerner, 2006; Moon, 2011; Gans and Murray, 2012) by examining, in particular, effects of disclosures *before* the creation of a final innovation.

[3] Other considerations may include allocative interests or fairness and minimizing deadweight losses. Here, however, we emphasize dynamic considerations of technical change and innovation, given our focus here on innovation processes and that such considerations are most central to advancing welfare in society (Solow, 1957).

[4] This presumes, for example, that lower appropriability is not somehow outweighed by benefits of wider adoption and reuse by competing innovators, as with say network effects (e.g., Cusumano et al., 1992), powerful complementarities

view that the absence of patents (a final disclosure mechanism) might decrease incentives (Schankerman, 1998), which enjoys at least some empirical support (Belleflamme, 2006). Note, this lower contractibility adds to what might already be depressed incentives under any form of reuse – whether with intermediate *or* final disclosure, as dividing payoffs between upstream and downstream innovators can itself harm incentives (Scotchmer, 1991, 2004; Green and Scotchmer, 1995).

Implications for follow-on reuse between final and intermediate disclosure are more straightforward. Intermediate disclosure – by definition – creates the opportunity for earlier, more frequent and wider ranging disclosures, with fewer restrictions on reuse than does final disclosure.[5] It immediately follows that intermediate disclosure enables greater reuse of a given upstream innovation, all else being equal.[6]

As straightforward as the logic of the "reuse" half of the "incentives-versus-reuse" tradeoff might be – i.e., all else being equal, removing obstacles to disclosure indeed enables greater reuse – it remains difficult to directly observe relevant all-else-being-equal empirical comparisons to illustrate this point. Nonetheless, we can begin to appreciate the central role of follow-on innovation and reuse under intermediate disclosure simply by referring to the numerous studies documenting their abundancy in contexts such as open source software projects (e.g.: von Krogh et al., 2003), communities of inventors (e.g. Meyer, 2013), user innovator groups (e.g., von Hippel, 2005) and other intermediate disclosure regimes.

Although follow-on reuse is often taken as a matter of fact in research focused on intermediate disclosure regimes, innovators' incentives to enter, participate and exert effort in development is taken as a "puzzle" given lower appropriability conditions (e.g., Bonaccorsi and Rossi, 2003; Lerner and Tirole, 2002) – consistent again with the incentives-versus-reuse tradeoff. In many cases, the puzzle is clarified by pointing to some form of compensating mechanism that countervails any productivity losses from depressed incentives. For example, enlisting "many eyeballs" (Raymond, 1999) through globally distributed reach and highly modular work streams can offset lost incentives in open source software projects. Lower contractibility and payoffs specifically related to disclosures and reuse might also be offset by enlisting sources of motivation not tied to transfers and reuse, such as intrinsic, pro-social, own-use and learning-by-doing motivations (Lakhani and Wolf, 2005; Roberts et al., 2006). Further, research on open source and user innovation routinely refer to populations of those innovators as "communities" (Franke and Shah, 2003; von Hippel and von Krogh, 2003; West and Lakhani, 2008; O'Mahony and Lakhani, 2011), implying norms and socialization might compensate by producing cooperative, reciprocating interactions rather than competitive ones. In some cases, intermediate disclosure regimes still involve attempts to tie payoffs to encourage effort and disclosure, at least for small levels of effort. For example, Wikipedia encourages "edits" (contributions and simultaneous disclosure) via the motivations of participants to signal their expertise to a wider audience of users (Zhang and Zhu, 2011). The small size of individual intermediate, partial contributions might also relieve the need for herculean efforts by any one individual.

Beyond papers focusing on patterns associated with particular disclosure policies, a small stream of recent papers attempts to compare follow-on reuse when disclosure policies vary within a given innovation system. None precisely relates to intermediate versus final disclosure, but nonetheless offers some corroborating insight. For example, several papers have begun to investigate effects of patents, a final disclosure mechanism, on follow-on reuse (relative to no patents, at all). The arguments presented here would predict that patents, as a final disclosure regime, should retard disclosure and reuse of a given upstream innovation all else being equal, relative to rates of reuse in intermediate disclosure. However, rather than simply finding patents generate little reuse, these studies fail to find any evidence whatsoever that patents produce an increase of reuse at all – and most data analyzed thus far suggests a *decrease* (Huang and Murray, 2009; Murray and Stern, 2007; Williams, 2013; Sampat and Williams, 2014).[7] Thus, rather than promoting reuse via property rights and a "market for ideas" (Kitch, 1977; Arora et al., 2004) at least these existing comparisons suggest imperfections and transaction costs of patents (Heller and Eisenberg, 1998; Anand and Khanna, 2000; Bessen, 2004) that lead them to support fewer disclosures than even our theory would allow for. More broadly, these results corroborate our contention that hindrances and conditions placed on disclosure can have large negative implications for on-going reuse.

Perhaps works that are closest to our reuse "half" of the incentives-versus-reuse tradeoff, are those by Furman and Stern (2011) and Boudreau (2010). Furman and Stern (2011) show that the establishment of biological resource centers as an infrastructure and contracting framework to grant access to research materials increases reuse by wide margins. Boudreau (2010) analogously shows that both granting access to and devolving control over upstream operating system platforms in personal digital assistants and smartphones accelerates reuse in downstream development sizably increases building of downstream products on top of those platforms.

We summarize the preceding arguments in the following prediction.

> ***Prediction* 1** ("Incentives-versus-Reuse" Tradeoff): Implementing an intermediate rather than final disclosure policy leads to lower incentives but greater follow-on reuse.

and increasing returns in knowledge recombination(Weitzman, 1998; Bessen and Maskin, 2009; Belenzon, 2012) or the establishment of a cooperative reciprocating or "sharing" equilibrium among those disclosing and recombining knowledge (e.g., Allen, 1983).

[5] Note that these arguments imply that contractibility is "good" for upstream innovators but "bad" for downstream reuse. This might seem a departure from Coasian reasoning, where the assignment of property rights and ability to write contracts can assure efficient trades. However, even if a perfect contract for the transfer and reuse *could* be written, the downstream innovator might not be able to commit to the contract without first engaging in trial-and-error experimentation to assess the value of an upstream input. Thus, the downstream innovator would need a means of accessing this knowledge prior to assessing its value. Alternatively, under less than ideal Coasian conditions, this contention (of good for upstream incentives and bad for downstream reuse) simply requires that imperfect contracts go some measure towards guaranteeing upstream innovators' payoffs, while adding transaction costs in downstream reuse relative to what would be the case in a less heavily contracted environment.

[6] The all-else-being-equal qualification is rather important here, as intermediate disclosure could affect incentives in a way that determines whether a given upstream innovation appears in the first place. It should also affect downstream innovator's incentives to invest in "absorbing" and learning upstream knowledge and technology. Intermediate disclosure could also shape coordination and search costs. For example, a published academic article (a final disclosure) comes with the validation, certification and screening of the peer review process. The selection process of journals and the full development and articulation of contributions in an article also help sort and categorize its content for other scholars to comprehend and situate within the torrent of academic research output.

[7] Williams (2013) is exceptional in finding zero effect of patents on disclosure. Her interpretation is that given there is zero effect on disclosure, the effect of patents on reuse can be disregarded in efficiency assessments of patents. Our theory and arguments instead would suggest that the absence of an effect on reuse should instead be regarded as cause to question to efficacy of the disclosure mechanism in patents. It should be noted too that Galasso and Schankerman (2013) find considerable variation from industry to industry, even in their relatively selected subsample of inventions that are invalidated, suggesting considerable need for caution in interpreting the existing evidence.

3.2. Disclosure policies and search for solutions

The earlier incentives-versus-reuse tradeoff draws on literatures that implicitly conceptualize the innovation process as a "production function" where upstream knowledge or technology serves as an "input" along with effort and investment in determining the level of downstream innovation that results. In this section, we consider that innovation performance is also shaped by the the search for problem-solving approach. We develop predictions in relation to disclosure policies.

3.2.1. Alternative approaches to solving problems

Innovation problems that involve making large numbers of interacting decisions are, by their nature, complex and uncertain – and therefore require *search* to find a solution (Simon, 1962). For example, in aeronautical engineering, landing gear design involves sets of decisions related to a great many interrelated parameters (e.g., the number and configuration of wheels, gear design, retraction and extension method, etc.) that trade off various performance criteria (e.g., drag, weight, cost, maintenance, reliability, etc.). Analog of such complex innovation and problem-solving exist in science, technology, software, artistic composition, etc.

In principle, there may be multiple solutions that meet some criteria – although perhaps trading off aspects of performance in different ways. Where the individual decisions (around particular parameters of a solution) are complementary and "go together" (Rosenberg, 1982) altogether different solution approaches can exist (e.g., a fixed landing gear goes together with the choice of aerodynamic covers for wheels, whereas a retractable landing gear goes together with modifications to the fuselage). Innovation thus tends to proceed as a search for *approaches* and then as a search for optimal solutions within a given approach. (This is akin to search on different "hills" within a "rugged landscape" (Kauffmann, 1993; Levinthal, 1997), or "exploration versus exploitation" (March, 1991)). In the history of aeronautical design, landing gear design proceeded in at least four distinct and parallel approaches until finally the retractable landing became the preferred approach (Vincenti, 1994). Landing gear design proceeds today within that same basic approach, with continuing incremental improvements.[8] This notion of search is explored in a range of research traditions beyond just innovation, including artificial intelligence, psychology, biology, evolution, organizational learning and others (*e.g.*, Cyert and March, 1963; Newell and Simon, 1972; Simon, 1962; March, 1991; Levinthal, 1997).

To the extent that innovation and problem solving is shrouded in uncertainty, individual problem solvers may initiate search according to their own initial stock of knowledge and beliefs (Rosenkopf and Nerkar, 2001; Rosenkopf and Almeida, 2003). Trial-and-error experimentation then provides feedback and insight – learning – accumulating upon an innovator's initial stock of knowledge. Given constraints of uncertainty and bounded rationality, the choice of search direction can be influenced by some combination of heuristics (e.g. Nelson and Winter, 1982), theoretical understandings (Fleming and Sorenson, 2004), analogies from comparable situations (Gavetti et al., 2005) and the problem solver's own initial endowment of knowledge and experience. In addition, problem solvers may not be entirely isolated, autonomous and independent of one another. They may and often are, searching whilst others do the same. This provides an opportunity to observe the activities and outcomes of the experimentation of other problem solvers. Depending on the payoff structure, observing others' action can also influence expected returns of making investments in given directions.

3.2.2. Disclosure and innovative search

The earlier incentives-versus-reuse tradeoff, on its own, suggests that the higher incentives of final disclosure might be associated with higher experimentation – inasmuch as higher incentives and effort translate to greater search efforts. Here we argue that disclosure policies should more fundamentally transform patterns of search.

There will be a greater degree of independence in the choice of search approach across innovators under final disclosure, where a steady stream of intermediate updating and observation of others' actions and choices is not possible. At the population level, this may imply some degree of "parallel" or uncorrelated search, inasmuch as individuals maintain some level of ignorance regarding outcomes of each other's trial-and-error search process (Nelson, 1961; Boudreau et al., 2011).

In the case of intermediate disclosures, a steadier stream of updates has the potential to result in far more "coordinated" or correlated responses, if individual problem solvers can observe and respond systematically to their own experimentation outcomes and to those of others. Intermediate disclosure, not only increases the immediacy and extent of transfers and reuse, but also telegraphs information about the existence and potential of alternative approaches. More informed and coordinated searching could potentially produce deliberately *differentiated* search paths, as when competing innovators expect higher returns from staking out new ground than from engaging in overlapping experiments (cf. Murray et al., 2009; Acemoglu, 2012). Following this argument, it is possible that a coordinated search could lead to greater diversity in solution approaches than that generated by independent experimentation. However, a countervailing argument, is that converging on established solution approaches economizes search costs and reduces uncertainty – creating incentives to replicate and incrementally extend what has been done. Following the weight of historical evidence – indicating a general tendency to convergent trajectories, dominant designs and the emergence of scientific paradigms (e.g., Kuhn, 1962; Utterback and Abernathy, 1975; Dosi, 1982), we make the following prediction:

> ***Prediction2*** ("Independent-versus-Convergent Search"): Implementing an intermediate rather than final disclosure policy leads innovators to tend to converge towards successful solution approaches and to engage in a lesser degree of independent experimentation.

4. Experimental design

To study the questions raised in the earlier discussion we implement a synthetic innovation system in a "Petri dish," applying different disclosure policies in different treatments. At the same time, we hold constant the problem addressed and the composition and number of participants working under different treatments so as to infer *ceteris paribus* effects of disclosure policies.

The research question and our objectives place relatively high restrictions on the research context. We needed an exceptionally highly controlled environment that would enable us to precisely implement distinct disclosure policies while controlling all other aspects of the institutional context and observing relevant micro-measures (i.e., effort, problem-solving performance and the particular solution approaches pursued).The experiment was carried out over two weeks on the online software development platform of TopCoder, the leading platform for developing software and algorithmic solutions as a series of contests. The platform has existing communications and payment systems and a membership

[8] Bijker et al. (1987) document multiple simultaneous approaches in a range of technologies including bicycles, synthetic dye chemistry and ship-building.

of elite software and algorithm problem-solvers. We worked closely with TopCoder executives and technologists to modify the platform to implement the features of the experimental design, the particulars of which we describe here.

4.1. Sign-up phase and random assignment to independent groups

Subjects were recruited from the TopCoder platform's existing membership of software and algorithm programmers, the experiment being included as part of the regular stream of listed "challenges" members can sign up for and participate in. The posting indicated that the challenge would feature an algorithmic optimization solution related to genomics, the solution to which was sought by Harvard Medical School (from which the problem had been sourced, see Section 4.2), that the total prize pool would be $6000.00 and that the exercise was also being used for research purposes. The call for participation did not describe what particular problem would be solved.

In response to our call for participation, 733 TopCoder members signed up for the experiment. Roughly half (44%) were professionals, the remainder were students. Participants from India (20%), the United States (16%), China (9%) and Russia (9%) accounted for more than half of a pool that represented 69 countries.

Our primary goal in the experiment was to observe differences not only across individuals working under different disclosure policies, but also differences across the groups, as a whole. For example, patterns of learning and innovation in each group is a collective outcome, at least as much as it is can be regarded at the individual level. In this sense, it was important to design an environment – synthetic as it might be – that would represent a *population* of creative problem-solvers could, in the presence of a given institutional framework, might elect to participate in the innovation process or otherwise devote their attention to outside options (much like the case in our real-world innovation systems). This, however, implies the need to create large comparison groups of prospective entrants, to enable us to observe entry, non-entry and the consequences of interactions among individual subjects who do enter and actively participate.

With these points in mind, we constructed a minimum – only two – main experimental comparison groups to maximize their size: the "Intermediate Disclosure" and "Final Disclosure" regimes. A cost of large groups is a loss of replication, with just one trial per treatment. As a means of providing greater validation to the results, we constructed a third, supplementary regime to better assure that our main comparison groups were not somehow eccentric. In this "Mixed" regime, intermediate disclosure was permitted in the second but not the first week of the experiment. In reporting results, we focus on the main comparison groups and return to the Mixed regime to find they corroborate main results.

In creating three equally-sized (Intermediate Disclosure was assigned 245 treatments; Final Disclosure and Mixed treatments were each assigned 244 subjects), independent groups of similar composition, we randomized, but simultaneously matched on skills by rank-ordering participants according to skill-ratings and then randomly assigned descending triplets of roughly equal skill ratings. We are able to observe skills, as the TopCoder platform provides participants' skill ratings, formulated via an Elo-based system (Maas and Wagenmakers, 2005). This rating estimates skill on the basis of historical performance in similar algorithmic problem-solving exercises.[9]

4.2. Problem-solving and development phase

All development and interaction occurred on the online platform. Once assignment of individuals was completed and problem-solving was set to begin, a series of information was immediately revealed. Subjects were given the problem statement, together with a description of how solution accuracy and speed would translate into quantitive solution scores. At the same time, the identities ("handles") of all other subjects in the same treatment group were revealed.

4.2.1. The problem

The innovation problem was to develop *de novo* solutions in computer code to a problem sourced from researchers at Harvard Medical School, specifically, to design a bioinformatics algorithm to compare and annotate a large series of genomic sequences. The problem involved processing large amounts of data, accurately annotating the sequencing while minimizing transcription errors, solving within constrained computational resources (utilizing only an off-the-shelf computer) and minimizing the amount of time. A detailed description of the scientific features of the problem and scoring of solutions is described by Lakhani et al. (2013).

The problem we selected sits at the intersection of software development, mathematics, computer science and biology, is non-trivial and challenging and is a sort of computational optimization problem that involves iterative solution development and ongoing incremental gains rather than a final analytical solution that is either correct or incorrect. The focus on algorithm development enables us to treat intermediate solutions as primary inputs to subsequent development within a trial-and-error learning process. Working in digital format also enables solutions to be codified and recorded in computer instructions and evaluated by an automated system. The algorithmic setting also enabled us to devise a common, automated and precise measure of quality. Such advantages would not be possible were we to use a non-digital context. Further, the specific problem is highly salient in the scientific literature, having first been addressed when gene sequencing got underway (Altschul et al., 1990). More generally, the problem is also representative of complex, data-intensive numerical optimization problems.

4.2.2. Work environment, scoring and information

Subjects worked through the platform's "heads-up" interface screen on which development could be performed in a range of computer languages. Subjects received direct feedback by submitting their algorithm designs to the platform for assessment by the automated test suite and observing solution scores. (It was not possible to receive direct feedback on the quality of submissions "off line.") The main task required subjects to write code to maximize accuracy and minimize time in identifying the originating gene segments that formed a particular genomic sequence. Each code submission had to evaluate 100,000 genomic sequences with the quality score determined as a linear combination of speed and accuracy. Although the explicit scoring model was shared with participants, practically speaking, it was only through submitting intermediate working solutions to the platform in trial-and-error fashion that participants could determine their scores and whether improvements could be made.

4.2.3. Disclosure regimes

The differences of central interest are those related to the disclosure policies that were implemented across the independent

[9] The average participant had 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 European Go Federation.

groups. Subjects were informed of procedures within their respective regimes, but not of procedures of other groups or that there would be differences across groups.

4.2.4. Final disclosure

Final Disclosure is the simplest case. No communications or sharing facilities were enabled on the platform for this treatment. Individuals were also instructed that interacting with other competitors off the platform would result in immediate disqualification of all involved. Participants simply worked on their own and submitted their solutions for scoring. Although solutions were not disclosed, we followed TopCoder's insistence that the system display scores and rank ordering of participants throughout the experiment.[10]

4.2.5. Intermediate disclosure

In Intermediate Disclosure all solutions submitted by subjects in the trial-and-error development process were immediately disclosed to other participants in the same treatment group. This was done via the same main "heads-up" display interface used to conduct software development. All submitted intermediate solutions, identified by score, submitter and time of submission, were listed and available to participants in their entirety (i.e., source code) by simply clicking on the relevant entry in the list. The implemented system of disclosure should thus be understood as relatively simple and frictionless.

4.2.6. Mixed regime

In the Mixed regime, intended to supplement our main comparisons, the first week followed the rules and procedures of Final Disclosure (i.e., no disclosures), the second week followed the rules and procedures of Intermediate Disclosure (i.e., all solutions from the first week was revealed along with any subsequent solutions during the second week).

4.3. Payoffs and rewards phase

Our primary interest in this experiment is to examine effects of variation in disclosure policies, nonetheless we define a fixed underlying institutional design held constant across treatments. Particularly relevant is, of course, the payoff structure.

4.3.1. Payoffs in final disclosure

Payoffs were tied to quality of the solutions developed, based on the precise quantitative scoring enabled by platform submissions. In general, payoffs to an innovation problem can, for example, go to the top performing solutions, as in winner-take-all outcomes. Such is the case where only the very best solution is desired and there is no need for variety of solutions or solution approaches in addressing a particular problem. However, more generally, we often see some returns to more than just one solution for a given problem (e.g., competition in academia, industrial competition, prize contests, etc.). For this reason, we chose an arbitrarily small number (5) of prizes, monotonically varying in size with rank. (We were also encouraged to follow this design by TopCoder executives, on the basis of their experience.) In Final Disclosure, the top five positions were allocated a total of $1000 in payoffs ($500, $250, $125, $75 and $50) at the end of each of the two weeks, i.e., a total of $2000. (The particular amounts were chosen under the advice of TopCoder executives, given the nature of the problem and given our interest in eliciting wide participation. TopCoder executives also recommended we break up payments over two weeks, rather than confer rewards only at the end, as a means of maintaining engagement and participation across the entire process, enabling us to observe a longer period of experimentation and improvement.[11]) Specifically, payoffs were based upon the final, last submissions made by each subject, each week. Top ranking subjects were also publicly announced on the TopCoder website. Therefore, implicit, reputational payoffs accompanied monetary payoffs.

4.3.2. Payoffs in intermediate disclosure

In Intermediate Disclosure we created an environment and framework in which both upstream problem-solvers and downstream providers of superior final solutions were conferred payoffs – maintaining the same overall budget as in Final Disclosure ($1,000 per week). We chose an arbitrary division of payoffs between final solution providers and upstream innovators of one-half. (see Green and Scotchmer (1995) for a more complete discussion of the theoretical considerations in allocating shares.) Therefore, in Intermediate Disclosure, the top five ranked solution providers were allocated a total at the end of each week of $500 ($250, $125, $62.50, $37.50 and $25), akin to the Final Disclosure case but half the magnitudes. The remaining payoff budget was allocated to upstream innovators.

In submitting any given solution, downstream innovators were instructed to list solutions they examined (i.e., clicked on through the web interface) and somehow drew upon in improving their own solutions. This involved listing the name of creator of solution that was drawn upon and answering the question of what fraction (percentage) of the solution drew on the identified prior submission. The response of this question had no bearing on payoffs to the downstream innovator (in the instance they would go on to secure one of the top five positions). The sum totals of percentage responses for the identified upstream contributors to the top five ranked solutions where then used to generate a rank order of all upstream innovators. Cash payoffs were then conferred to these upstream innovators in the same magnitudes as the top final solution providers (i.e., $500, $250, $125, $75 and $50). This payoff structure therefore implements a framework in which rewards are allocated to high-performing upstream solutions. However, stipulations of downstream use including payoffs and rewards associated with reuse are minimal in a low contractibility environment.[12]

Fig. 1 summarizes timing and payoffs across the distinct comparison groups in the experiment.

5. Data and variables

The data set we analyze here was constructed from several data sources. First, the experimental setup provides us with observational data subjects' decision to participate and make at least one submission or not, number of submissions, quality of submissions. As a supplementary measure of effort exerted, apart from number of submissions, we also administered a survey after all coding activity was completed and before final results were

[10] This creates the possibility of updating and consequent decision-making throughout the contest. However, the experiment was not designed to derive inferences in relation to any dynamic adjustments. We focus on cross-sectional comparisons.

[11] An alternative design would entail a lump disclosure of all first-week solutions at the end of the first week. However, we elected to implement the simpler design as it did not bear on predictions in Section 3. There is clearly scope for examining more nuanced dynamics in future research.

[12] Note, here we do not generate low contractibility by varying the form of disclosures (see Section 2.3), but rather we directly implement low contractibility in how we define the experimental framework.

Table 2
Description of variables.

Variable	Unit of observation	Description
Participation	Subject	An indicator variable switched on for all subjects submitting at least one solution.
Submissions	Subject	Count of the number of solutions submitted by a subject over the course of the two-week experiment.
No. disclosures	Subject	Count of the number of past solutions for which the subject "clicked through" to see the complete code.
Submission quality	Solution submission	An automated precise measure of the quality of a each solution submission.
Programming language	Solution submission	A wide variety of programming languages were admissible. This field records the name of the language.
Solution approach	Solution submission	A 10-digit binary code describing which among 10 elemental techniques are used in a given solution submission.

Note. The data set relates to 733 subjects, of which 122 are submitters, submitting 654 solutions.

made public, asking participants to report the number of hours worked over the two-week period. The experimental setup also allowed us to record subjects' clicking through to examine disclosed code submissions of other participants under Intermediate Disclosure. Key variables in the analysis are described in Table 2. Descriptive statistics of variables are provided in the analysis to follow.

We also collected the software code for each submission. From this, we recorded the programming language utilized. Further, we hired three Ph.D. experts to uncover and systematically document the technical solution approaches used in each of the 654 submissions. The experts first reviewed the submissions to infer that across the body of solutions, subsets of distinct combinations of ten elemental computational techniques were used (Table 3, see Lakhani et al. (2013) for more details on the approaches developed in the experiment). Each solution was then coded according to whether each of these techniques were employed, leading each solution to be encoded as a 10 digit binary code. Across each of the submissions, the experts identified 56 distinct 10-digit combinations or solution approaches. Consistent with the innovation literature's identification of novel approaches as combinations of distinct knowledge sets (Fleming, 2001; Fleming and Sorenson, 2001), unique combinations of methods can be understood to represent distinct approaches to varying degrees, while there is still considerable variation in the particular implementation and quality across individual solutions within a given approach.

6. Results and analysis

Before proceeding to assessing evidence in relation to Predictions 1 and 2, Fig. 2 first provides a broad orienting overview of problem-solving patterns in the main comparison groups, showing submissions by quality, over time. The graphs also trace the maximal frontier (black) and moving averages (red) lines. Table 4 provides a broader overview of readily-observable cross-sectional differences between Intermediate Disclosure and Final Disclosure outcomes that we will discuss in greater detail in the discussion to follow.

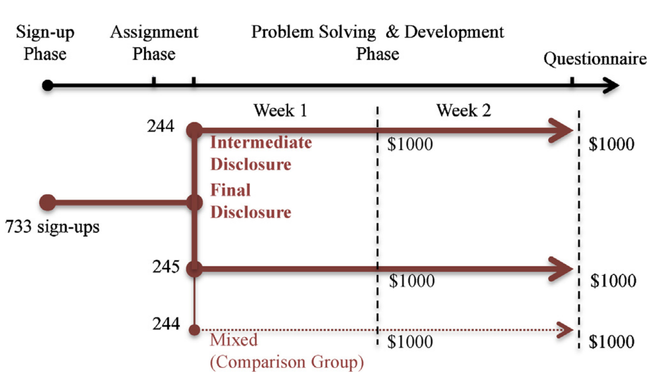

Fig. 1. Overview of experimental sequence.

6.1. Findings on the incentives-versus-reuse tradeoff

Results presented in this section relate to Prediction 1. We examine evidence in relation to either side of the tradeoff in turn.

6.1.1. Effort and incentives

Evidence of lower levels of participation and effort exerted under Intermediate Disclosure are consistent with our prediction of lower incentives under Intermediate Disclosure, as in Prediction 1. (See "'Entry' Participation and Effort" in Table 4.) The fraction of subjects choosing to "enter" and actively participate in problem-solving (i.e., subjects submitting at least one solution) was significantly – 26% – lower in Intermediate Disclosure than in Final Disclosure, with 14% entering and actively participating in Intermediate Disclosure versus 19% in Final Disclosure. Precision and significance of this difference increase when controlling for

Table 3
Elemental techniques used in solutions.

Method	Description
1	**Filtering by unmapped alignment scone (Hamming distance)**: Compare the query string against strings from sets A, B and/or C, trying various possible offsets.
2	**Filtering by comparing frequencies of hashed chunks:** For both the query string and strings from A, B and/or C. move a sliding window across the string and make a frequency table of the chunks that appear in the window, optionally after hashing the chunks. Select the best match(es) between the frequency table obtained from the query and those from the carpus.
3	**Dynamic programming:** Compute the actual Levanshtein distance between a portion of the query and strings from sets A, B and/or C.
4	**Dynamic programming extended to more than one section (A, E, C) at once:** Extend the dynamic programming Levenshtein distance computation to find the optimal edit distance between (a portion of) the query and all possible A + B, B + C or A + B + C combinations.
5	**Bit optimizations:** Use bitwise arithmetic to operate or multiple characters at a time.
6	**SEE optimizations:** Use Streaming SIMB Extensions (a CPU instruction set enabling single-instruction multiple-data (SIMD) parallelization) to process up to 16 characters or strings at once.
7	**Refinement of choices after finding initial solution:** As a post-processing step, hold two of the three selections fixed and reoptimize the third.
8	**Fast approximation of edit distance in well-matched regions:** Use restricted dynamic programming. Hamming distance or variants thereof to speed up the cumputation.
9	**Precomputation of statistics on the string corpus:** Perform offline analysis of the provided sets A, B and C, and use the data obtained for decision making in the algorithm.
10	**Explicitly prefer shorter B strings:** In heuristic approaches, give bonuses to shorter strings from set B (which empirically have greater likelihood of producing high scores).

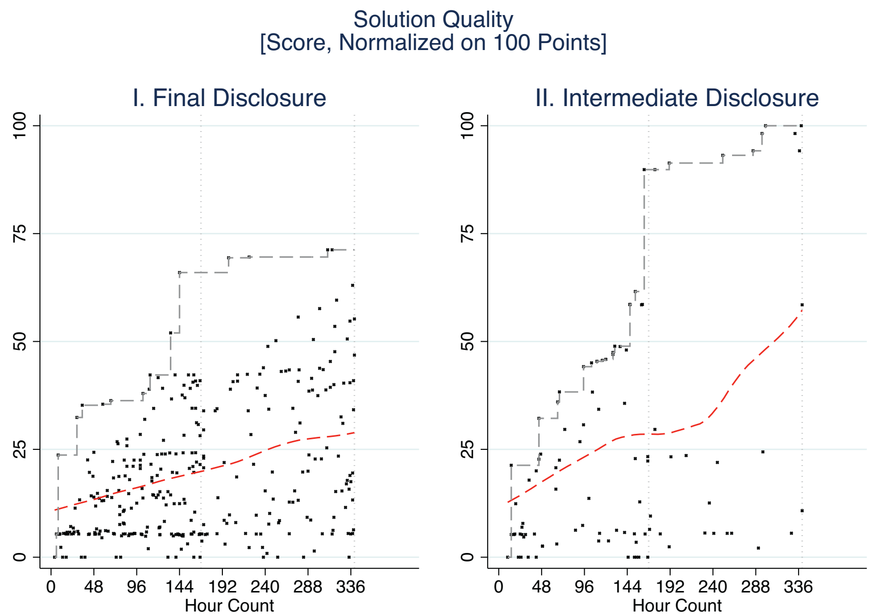

Fig. 2. Overview of solution submissions over time, by disclosure policy (submissions, fitted mean and maximal envelope shown).

Table 4
Overview of differences in outcomes across final disclosure and intermediate disclosure.

	Final disclosure		Intermediate disclosure	
	Mean	S.E.	Mean	S.E.
No. subjects assigned:	245		244	
"Entry", participation and effort				
No. active participants/submitters	46		33	
Prob{Submitting}	.19	(.01)	.14	(.01)
No. submissions	319		99	
Submissions/active participant	6.93	(.42)	3.90	(.14)
Examinations of intermediate solutions				
No. examinations of intermediate solutions	N/A		1359	
No. examinations by submitters			1024	
No. submitters examining			30	
Prob{Submitter examining}			.91	(.01)
No. solutions examined per examining submitter			34.13	(1.79)
No. examinations by non-submitters			335	
No. non-submitters examining			46	
Prob{Non-submitter examining}			.22	(.01)
No. solutions examined per examining non-submitter			7.28	(.5)
Solution approaches				
No. solution approaches	27		19	
Unique (to Submitter) approaches per submitter	1.96	(.07)	1.67	(.06)
Unique (to Group) approaches per submitter	.59	(.07)	.58	(.06)
Unique approaches per submission	.08	(.00)	.19	(.01)
Problem-solving performance [final scores]				
Max	71		100	
q90	47		94	
q75	24		38	
Median	13		11	
q25	5		5	
q10	5		2	
Min	0		0	

skill levels and other subject covariates, without altering the point estimate of the difference (not reported).[13]

Those who did participate under Intermediate Disclosure also exerted lower effort, despite facing fewer other active participants as competitors. The number of submissions per participant was 56% lower in Intermediate Disclosure than in Final Disclosure (3.9 versus 6.9 solution submissions, a difference significant at $p = 1\%$). Our other measure of effort – the number of self-reported hours worked over the two weeks (Section 5) – is also lower in Intermediate Disclosure – 10.0 versus 14.1 hours, a difference significant at $p = 1\%$, based on a 60% response rate. (Precision and significance of these differences again become greater still when controlling for skill level while not changing estimated magnitude of differences.) Therefore apart from considerably more participants

[13] The data suggests results are driven by "treatment effects" and behaviors rather than "selection effects" and composition. For example, we found no statistical evidence of differences in distributions of skills, industry, areas of technical interest, age and self-reported source of motivation for participating on the platform.

entering to engage in problem-solving under Final Disclosure, each active participant also exerts considerably greater effort. In total, Intermediate disclosure produces 99 solution submissions in relation to 319 produced under Final Disclosures.[14] Taken together,[15] we interpret patterns of lower entry and active participation, lower hours worked and lower numbers of submissions as consistent with lower incentives associated with Intermediate Disclosure.

6.1.2. Reuse of intermediate solutions

Intermediate Disclosure, by definition, enabled intermediate disclosures and reuse whereas Final Disclosure did not. It necessarily follows that the reuse half of the "incentives-versus-reuse" tradeoff holds (Section 3.1).

Beyond simply being greater than zero, we find that disclosures and reuse in this environment were widespread and frequent. In total, 30 of the 33 active participants submitting at least one solution in Intermediate Disclosure examined a total of 1,024 intermediate solutions, or 34.1 on average (std. dev. = 28.0). This is consistent with the use of a relatively frictionless ("click-through") system, an ease of reviewing these intermediate solutions, the focus of all parties on the same problem and thus, an expectation of returns to reviewing intermediate solutions. Early and highest-scoring, solutions elicited the greatest number of examinations. Incidences steadily declined over the two-week duration, with four-fifths of examinations of prior solutions occurring in the first week of the experiment.

It was also the case that many subjects who did not eventually enter and participate also examined solutions, 44, examined intermediate solutions of others. However they examined many fewer 7.3 on average (std. dev. = 8.5). This might be explained in relation to curiousity or an interest in learning. It is also plausible that some of these non-submitters might have decided not to submit upon having reviewed others' solutions.

6.2. Findings on differences in the innovation search process

Results presented in this section relate to Prediction 2.

6.2.1. Performance patterns and trajectories

As a first approach to inferring the nature of search in either regime, we contrast the performance trajectories of individual submitters, as in Fig. 3, by connecting the dots representing submissions by the same submitter. Perhaps more than any other figure or table in the paper, this figure reveals the workings of our two main predictions at once in influencing innovation outcomes. (Participants who did not submit solutions do not appear.)

Patterns under Final Disclosure, in the left panel of Fig. 3, are inherently the most difficult to read. There are more numerous lines and dots and seemingly more erratic, less regular patterns. The greater number of lines and dots follows the earlier discussion of higher incentives, participation and effort (Section 6.1). Beyond a slight, if not entirely general tendency for individual performance trajectories to increase over time, some start high, others low, at times declining, at times increasing. There is not clear indication of correlated or coincident perturbations across submitters. This results in a high frequency of solutions distributed relatively evenly across the performance spectrum overall and in every time period (Fig. 3). These patterns are consistent with independent trial-and-error learning and experimentation occurring under Final Disclosure, as theorized in Section 3.2.

Patterns under Intermediate Disclosure, in the right panel of Fig. 3, starkly contrast with those of Final Disclosure. Differences begin with there simply being fewer trajectories and fewer individual submissions. Rather than the up-and-down trajectories of Final Disclosure, we observe laminar, smooth patterns, ascending *together*. Individuals' trajectories (save for those of a handful of low scoring outliers) also cluster on the maximal performance envelope and increasingly do so over time. These patterns are consistent with greater coordinated patterns of learning, experimentation and advance across subjects in a collective process of cumulative innovation. In this, the shape of trajectories suggest also a tendency towards convergence rather than differentiation.

Thus, these patterns documented in Fig. 3 are consistent with Prediction 2.

6.2.2. Diversity of solution approaches

An added perspective unto Prediction 2 is provided by data on solution approaches (i.e., combinations of the 10 elemental techniques, see Section 5). These data affirm the earlier suggestion of a greater tendency to coordination in the form of convergence rather than divergence in the case of Intermediate Disclosure.

Fewer solution approaches were tried by the overall group of submitters in Intermediate Disclosure. A total of 19 unique solution approaches were developed under Intermediate Disclosure whereas 27 unique solution approaches were developed under Final Disclosure (i.e., 42% fewer in Intermediate Disclosure). Fig. 4 presents the accumulation of distinct solution approaches over time. The lines never cross, indicating there were always a greater number of approaches attempted in Final Disclosure throughout the entire exercise. In Intermediate Disclosure, apart from fewer overall solutions (99 versus 319) and fewer solution approaches (19 versus 27), there were fewer programming languages used. In Intermediate Disclosure, three languages (C#, C++ and Java) were used; in Final Disclosure these languages and two more were used (Python and Visual Basic).

It remains possible, nonetheless, that lower levels of experimentation and diversity are simply the result of lower incentives and effort exerted under Intermediate disclosure. It is difficult to entirely rule this possibility out and we should expect this played some sort of role. However, there is evidence consistent with convergence in directions of experimentation. Submitters under Intermediate Disclosure themselves, as individuals experimented across 15% fewer solution approaches than did those in Final Disclosure (1.67 versus 1.96, on average), as in Table 4 under "Solution Approaches". Given fewer submissions and solution approaches overall in Intermediate Disclosure, we might expect even randomly selected approaches to be less likely to overlap – and to be unique in relation to those used by the wider group. Any deliberate attempts to differentiate or even to engage in independent experimentation might then lead to a greater number of unique approaches pursued per submitter under Intermediate Disclosure. However, we find no such evidence. Unique solution approaches (relative to the group) per submitter are virtually identical (0.58 and 0.59, in Intermediate and Final Disclosure.)

[14] We should qualify this statement by noting that notwithstanding the large differential, even 99 solutions should be regarded as a (very) high number in relation to the level of usual level of attention and publishing around comparable algorithmic problems in genomics in recent decades.

[15] The measure of number of submissions has the appeal of being observational data, whereas hours worked is closest to a direct measure of effort level but is self-reported. The 29% drop in number of hours worked, however, is less susceptible to other factors' influence than is the 56% drop in numbers of submissions in Intermediate Disclosure. For example, subjects might delay submissions, foregoing a measure of trial-and-error feedback, to limit undesired disclosures. Public disclosure of others' solutions information regarding the efficacy of different approaches might also reduce the information value of making one's own submissions, while increasing the information value to simply studying prior solutions. The sharing of information regarding past solutions even to non-entrants might also have shaped decisions to enter either upward or downward, if this further informed their beliefs regarding their possible success. Alternatively, the reuse of prior solutions could facilitate and hasten submissions.

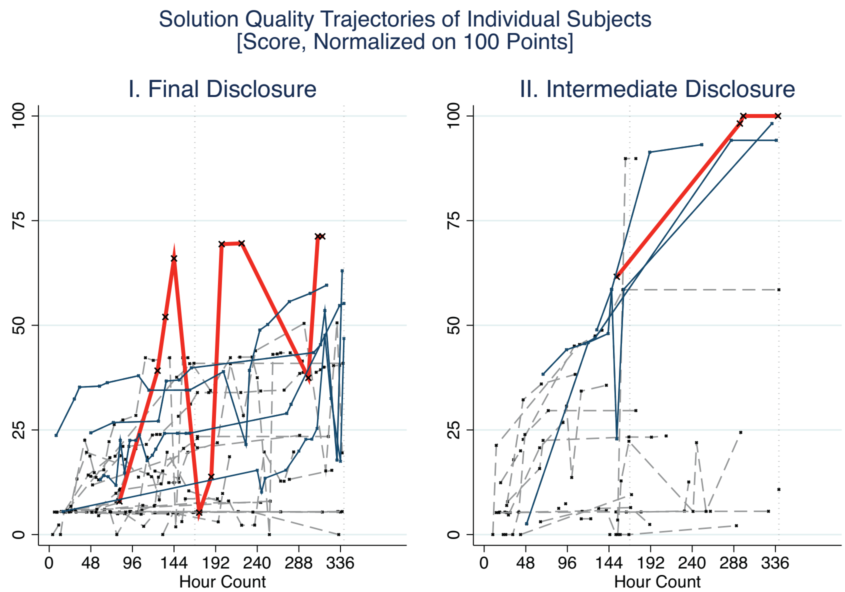

Fig. 3. Subjects' performance trajectories (highlighting first through fifth ranked).

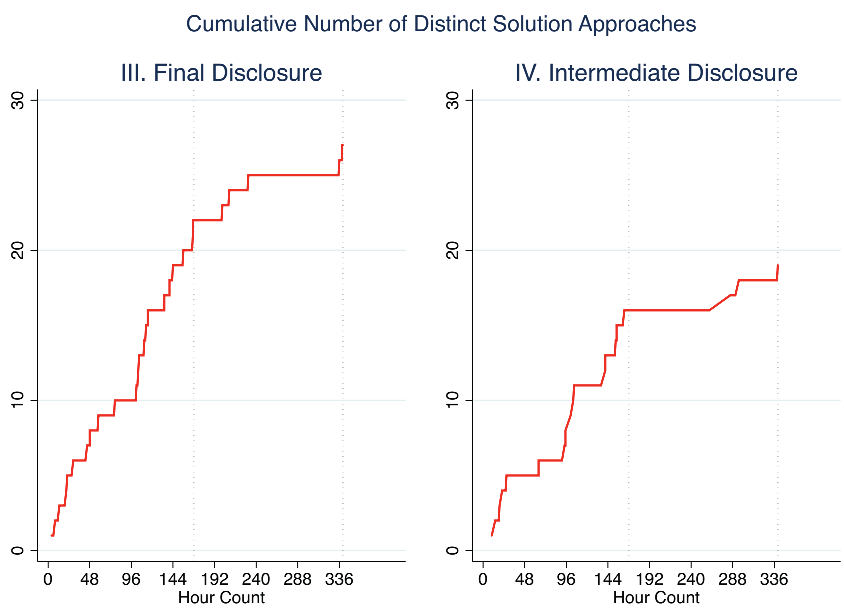

Fig. 4. Solutions by approach (i.e., unique combination of techniques).

Further suggestion of convergent coordinated approaches comes from direct measures of concentration of solution submissions across approaches. The Herfindahl measure of concentration of shares is 52% higher in Intermediate Disclosure (0.149) than in Final Disclosure (0.0986).[16] Beyond being concentrated, submissions in Intermediate Disclosure were concentrated on just "high-potential" approaches. We rank ordered each of the 54 solution approaches in the entire experiment by their "potential," based on the top score achieved within each approach and found that *every* solution in Intermediate Disclosure employed an approach that was above median.

6.3. Replication

Our experimental design minimized replication in order to maximize group size. Thus, we observe results in single trials of Intermediate Disclosure and Final Disclosure. This leaves the possibility that random eccentric outcomes or "butterfly effects" could have still somehow emerged and distorted population-level patterns in some way.[17] Some level of assurance already comes from

[16] The Herfindahl measure of concentration is the sum of squared shares. Therefore, a higher Herfindahl measure indicates higher concentration. We calculate the share of submissions for each approach within each treatment by dividing the number of submissions using a particular solution approach, by the total number of submissions in the treatment.

[17] Single-trial comparisons (i.e., where a stream of outcomes resulting from one policy is compared to a stream of outcomes from another) remain the norm in the literatures making before-and-after or differences-in-differences comparisons on effects of innovation policies.

Table 5
Regularity of results in relation to mixed comparison group.

	Final disclosure Mean	Mixed comparison group Mean	Intermediate disclosure Mean
No. subjects assigned	245	244	244
No. active participants	46	43	33
No. submissions	319	236	99
No. examinations	0	654	1359
No. solution approaches	27	25	19
Max	71	83	100
q90	47	61	94
q75	24	39	38
Median	13	24	11
q25	5	4	5
q10	5	5	2
Min	0	0	0

results conforming closely to *a priori* reasoned theory and predictions. Further, despite single trials at the population level, we have considerable replication at the subject level, where there are no obvious signs of unduly influential or rogue outlier data points.

In addition, we sought a minimal level of added empirical validation by running a Mixed treatment (Section 4.2.3). This is not replication in the sense of running multiple trails of Intermediate and Final Disclosure. It is nonetheless a basis for assessing consistency of outcomes with a single separate empirical benchmark trial. Table 5 reports key summary statistics for the Mixed regime, relative to the main comparison groups. The results suggest considerable regularity of reported patterns (Table 5). In the case of each of the main statistics reported earlier – number of participants, number of submissions per participant, number of examinations of intermediate solutions, numbers of solutions generated and maximum score achieved – the Mixed regime falls between Intermediate and Final Disclosure regimes. The precise distribution of scores (beyond maximum scores), given by quantiles of final scores, is less regular. Broadly speaking, these comparisons demonstrate considerable regularity of results, providing no indication of eccentric outcomes driving the earlier reported patterns.

7. Conclusion

This paper introduced an experimental framework for studying effects of disclosure policies on cumulative innovation, while contributing to a growing research interest in disclosures, transfers and "sharing" of knowledge and technology among innovators: "open innovation" of various sorts (e.g., West, 2003; von Hippel, 2005; Chesbrough et al., 2006; Laursen and Salter, 2006; von Krogh et al., 2003; Murray et al. 2009; Boudreau, 2010; Dahlander and Gann, 2010; Furman and Stern, 2011; Williams, 2013). Our goal here being to distinguish effects of intermediate versus final disclosure policies.

This paper contributes to this growing body of work – by first taking two steps back. The first step back we take is to observe that "open" innovation is hardly an isolated or exceptional phenomenon, if by "open" we mean that innovation takes place within a framework or system that deliberately enables transfers and reuse. We discussed and presented numerous examples (Table 1), to illustrate that the intended enablement of knowledge and technology transfers are a routine feature of most *every* innovation system – including those implemented by both public and private actors. This only stands to reason. Where innovators differ in their capabilities to recombine past innovations into new ones (and it is not always the originating upstream innovator who is superior in carrying this out) there will be innovative gains from transfers and reuse taking place. Designing (open) innovation systems therefore entails establishing frameworks in which productive transfers and exchanges – be they bilateral or multilateral – are feasibly implemented.

The second step back is to return to the longtime conception of the process of ongoing innovation as depending not just on transfers and reuse of knowledge and technologies *but also* on maintaining incentives (e.g., Romer, 1990; Green and Scotchmer, 1995). In this regard, this study departs from recent studies that focus on reuse and ongoing innovation patterns, without simultaneously considering how *ex post* reuse might affect *ex ante* incentives to develop the upstream innovation in the first place. We also depart from usual focus on a given institutional setting and innovation system in favor of considering approaches that distinguish altogether different systems, following intermediate disclosure versus final disclosure policies. Thus, we consider effects of disclosure *prior to the work on an innovation even being completed* rather than optimal length of patents and other timing issues that have been studied in the past.

Our work is somewhat analogous to pioneering econometric studies of naturally occurring contexts examining *ex post* reuse, particularly focused on effects of patents on on-going innovation (e.g., Murray et al., 2009; Galasso and Schankerman, 2013; Williams, 2013; Sampat and Williams, 2014). As patents are, in principle and by design,[18] intended to ease reuse through assigning property rights (Kitch, 1977; Arora et al., 2004), these patent citation studies can be interpretted as tests of the patent system's ability to deliver on this goal. None has yet found evidence of accelerated reuse of an innovation subsequent to it being patented. The results are thus consistent with transaction cost impediments to disclosures and reuse with patents (e.g., Heller and Eisenberg, 1998) (presuming there are indeed gains from trade/transfers in the contexts studied) – rather than addressing fundamental differences across systems as we do here. Our theory is consistent with but does not particularly strongly rely on these results. In terms of methods, our direct comparison of independent experimental groups under different disclosure treatments is also analogous to the comparison of small numbers of cases with and without patents in these studies. However, rather than compare cases with and without patents or investigating any one innovation system, our study sought to better understand differences across intermediate and final disclosure systems, an essential difference across a wide range of innovation systems. (The patent system, for example, is but one example of a final disclosure system.) Our work is also somewhat related to studies by Furman and Stern (2011) and Boudreau (2010), which document cases in which attempts to deliberately accelerate *ex post* reuse indeed did so.

7.1. Incentives-versus-reuse and patterns of "search"

We develop theory on intermediate versus final disclosure by drawing on insights from the economics of innovation (e.g., Green and Scotchmer, 1995; Bessen and Maskin, 2009; Murray et al., 2009) and a distinct tradition considering innovation as a complex "search" process (e.g., Nelson, 1961; Simon, 1962; Levinthal, 1997; Fleming, 2001) – showing that *both* perspectives are required to understand the most important, first-order determinants of innovation outcomes. A key starting point is our observation that intermediate disclosures reduce *contractibility* over transfers and reuse – where contractibility over innovations is already notoriously challenging under best of instances. Lost contractibility here refers specifically to the originating innovator's ability to uphold stipulations of reuse, including assuring their payoffs and

[18] Indeed, the word "patent" derives from Anglo-Norman "lettre patente," meaning "open letter."

rewards. By the same token, intermediate disclosure leads to earlier, more frequent and potentially wider ranging disclosures, with fewer stipulations and restrictions on reuse. Under usual contracting conditions surrounding innovation, these conditions imply an incentives-versus-reuse tradeoff, which was readily revealed in our experimental results. Intermediate disclosure led to 70% fewer solution submissions.

Apart from the incentives-versus-reuse tradeoff, our experimental analysis starkly revealed the importance of "search directions" and how systematic differences in choices of solution approaches across disclosure policies were essential to shaping innovation outcomes. Intermediate disclosures – if only by virtue of their timing – increase information and signaling in the innovation environment. As a result, choices of solution approaches across innovators are less independent, they are more "coordinated." In our experiment, coordination of decisions led to convergence, with 30% fewer solution approaches tried by those working under intermediate disclosure.

7.2. Comparative advantages of intermediate and final disclosure policies

The theoretical and experimental analysis should be understood as identifying simplest first-order tradeoffs and tensions created by different disclosure policies, necessarily abstracting from the many details of any one innovation system. The research design was particularly geared to documenting starkest differences in innovation outcomes on the basis of uncomplicated cross-sectional comparisons and with a minimum of econometric manipulation. Although a great many questions remain, the results begin to suggest the outlines of a "division of labor" between intermediate and final disclosure approaches, while highlighting limitations and challenges of each.

In our setting, intermediate disclosure promoted efficient reuse, coordination and convergence on a globally optimal solution with less entry and effort (i.e., lower costs) and higher performance. However, more generally and in a "rugged" landscape of possible solutions, we might be concerned that intermediate disclosure encourages path dependence and lock into a suboptimal solution approach, or leads incentives to evaporate. Such systems might therefore benefit from offsetting features of their design to countervail these weaknesses, as being directed – for example – to problems where the optimal solution approach is well known and wide experimentation is less useful and where returns to reuse are especially high. Alternatively, drawing on a wide and diverse pool of innovators less likely to fall into "groupthink" and to select on innovator types whose motivations are less dependent on contractibility of transfers and reuse. Inasmuch as intermediate disclosures imply smaller units of innovation output (e.g., edits, contributions, ideas, bug reports) many more individuals may be able to participate by making much smaller effort investments.

In theory and in our experiment, final disclosure promotes higher levels of entry and effort and independent experimentation. On the one hand, this generated wide diversity of approaches; on the other hand, this led to considerable effort devoted to suboptimal approaches and overall lesser learning and performance achieved. The overall empirical result of lower performance under final disclosure should hardly be regarded as general; tradeoffs should vary in importance according to the particularly prevailing structural conditions. Nonetheless, we might surmise that such systems might therefore also benefit from offsetting features of their design to countervail these weaknesses. This includes being devoted to conditions where wide diversity of experimentation is highly valued. Alternatively, if capabilities tend to concentrate and accumulate in individual innovators and there is little benefit from drawing on widely distributed contributions, then there may be higher returns to simply maximizing incentives of greatest experts, foregoing some degree of reuse.

Acknowledgements

We thank the Harvard Medical School and TopCoder for contribution of considerable support to this project, and especially note efforts of Ramy Arnaout, Eva Guinan, Lee Nadler, Jack Hughes, Rob Hughes, Mike Lydon, and Ira Heffan. We thank Po-Ru Loh, Hernàn Amiune and Xiaoshi Lu for technical evaluations of the data algorithms developed in the experiment. We also thank J.P. Benoit, Lee Branstetter, Wesley Cohen, Carliss Baldwin, Erik Brynjolfsson, Chaim Ferschtman, Francisco Gomes, Rebecca Henderson, Nicola Lacetera, Alan McCormack, Anita McGahan, Petra Moser, Ramana Nanda, Richard Nelson, Richard Stallman, Catherine Tucker, Keyvan Valkili, Feng Zhu and seminar participants at the at the Berkman Center at Harvard University, BYU-Utah Strategy Conference, London Business School, the National Bureau of Economic Research (NBER), the Open and User Innovation Workshop (2012), the Roundtable for Engineering Entrepreneurship Research (REER), Stanford University, Tel Aviv University, and the University of Michigan for helpful comments. Onal Vural provided excellent research assistance. 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 and the Research Program on the Economics of Knowledge Contribution and Distribution funded by the Sloan Foundation. Lakhani would like to acknowledge the support of the HBS Division of Research and Faculty Development, the NASA Tournament Lab and the MacArthur Foundation Network on Open Governance. A Google Faculty Research Grant supported both authors. All errors are our own.

References

Acemoglu, D., 2012. Diversity and technical progress. In: Lerner, J., Stern, S. (Eds.), The Rate and Direction of Inventive Activity Revisited. University of Chicago Press, Chicago, IL, Chapter 6.

Aghion, P., Dewatripont, M., Stein, J.C., 2008. Academic freedom, private-sector focus, and the process of innovation. RAND Journal of Economics 39 (3), 617–635.

Allen, R., 1983. Collective invention. Journal of Economic Behavior and Organization 4, 1–24.

Altschul, S., Gish, W., Miller, W., Myers, E., Lipman, D., 1990. Basic local alignment search tool. Journal of Molecular Biology 215, 403–410.

Anand, B.N., Khanna, T., 2000. The structure of licensing contracts. The Journal of Industrial Economics 48 (1), 103–135.

Arora, A., Fosfori, A., Gambardella, A., 2004. Markets for Technology: The Economics of Innovation and Corporate Strategy. MIT Press, Cambridge, MA.

Arrow, K., 1962. Economic welfare and the allocation of resources for innovation. In: Nelson, R. (Ed.), The Rate and Direction of Inventive Activity. Princeton University Press, Princeton.

Arrow, K.J., 1971. Essays in the Theory of Risk-Bearing. North-Holland Publishing Company, Amsterdam, The Netherlands, pp. 152.

Baldwin, C., Clark, K., 2000. Design Rules: The Power of Modularity. The MIT Press, Cambridge, MA.

Basalla, G., 1988. The Evolution of Technology. Cambridge University Press, Cambridge, UK.

Belenzon, S., 2012. Cumulative innovation and market value: evidence from patent citations. The Economic Journal 122 (559), 265–285.

Bessen, J., 2004. Holdup and licensing of cumulative innovations with private information. Economics Letters 82 (3), 321–326.

Bessen, J., Maskin, E., 2009. Sequential innovation, patents, and imitation. RAND Journal of Economics 40, 611–635.

Belleflamme, P., 2006. Patents and incentives to innovate: some theoretical and empirical economic evidence. Ethical Perspectives: Journal of the European Ethics Network 13 (2), 267–288.

Bijker, W.E., Hughes, T.P., Pinch, T.J., 1987. The Social Construction of Technological Systems: New Directions in the Sociology and History of Technology. MIT Press, Cambridge, MA.

Bikard, M., 2014. Hurdles to Invention Based on Academic Science: Evidence from 'Knowledge Twins'. http://ssrn.com/abstract=2333413

Bonaccorsi, A., Rossi, C., 2003. Why open source software can succeed. Research Policy 32 (7), 1243–1258.

Boudreau, K., 2010. Open platform strategies and innovation: granting access vs. devolving control. Management Science 56, 1849–1872.

Boudreau, K., 2012. Let a thousand flowers bloom? An early look at large numbers of software apps developers and patterns of innovation. Organization Science 23, 1409–1427.
Boudreau, K., Hagiu, A., 2009. Platforms rules: Multi-sided platforms as regulators. In: Gawer, A. (Ed.), Platforms, Markets and Innovation. Edward Elgar, London.
Boudreau, K., Lacetera, N., Lakhani, K.R., 2011. Incentives and problem uncertainty in innovation contests: an empirical analysis. Management Science 57, 843–863.
Bresnahan, T.F., Greenstein, S., 1999. Technological competition and the structure of the computer industry. The Journal of Industrial Economics 47 (1), 1–40.
Chesbrough, H., 2003. Open Innovation. Harvard Business School Press, Boston, MA.
Chesbrough, H., Vanhaverbeke, W., West, J. (Eds.), 2006. Open Innovation: Researching a New Paradigm. Oxford University Press.
Cohen, W.M., Nelson, R.R., Walsh, J.P., 2000. Protecting their intellectual assets: Appropriability conditions and why US firms patent (or not), NBER Working Paper No. 7552.
Contreras, J., 2011. Bermuda's legacy: policy, patents, and the design of the genome commons. Minnesota Journal of Law Science & Technology 12, 61–125.
Cusumano, M.A., Mylonadis, Y., Rosenbloom, R.S., 1992. Strategic maneuvering and mass-market dynamics: the triumph of VHS over Beta. Business History Review 66 (01), 51–94.
Cyert, R.M., March, J.G., 1963. Behavioral Theory of the Firm. Prentice-Hall, Englewood Cliffs, NJ.
Dahlander, L., Gann, D.M., 2010. How open is innovation? Research policy 39 (6), 699–709.
Dasgupta, P., David, P., 1994. Toward a new economics of science. Research Policy 23, 487–521.
Dosi, G., 1982. Technological paradigms and technological trajectories. Research Policy 11, 147–162.
Fauchart, E., von Hippel, E., 2008. Norms-based intellectual property systems: the case of French chefs. Organization Science 19, 187–201.
Fleming, L., 2001. Recombinant uncertainty in technological search. Management Science 47 (1), 117–132.
Fleming, L., Sorenson, O., 2001. Technology as a complex adaptive system: Evidence from patent data. Research Policy 30 (7), 1019–1039.
Fleming, L., Sorenson, O., 2004. Science as a map in technological search. Strategic Management Journal 25, 909–928.
Franke, N., Shah, S., 2003. How communities support innovative activities: an exploration of assistance and sharing among end-users. Research Policy 32 (1), 157–178.
Franzoni, C., Sauermann, H., 2014. Crowd science: The organization of scientific research in open collaborative projects. Research Policy 43 (1), 1–20.
Furman, J., Stern, S., 2011. Climbing atop the shoulders of giants: the impact of institutions on cumulative research. American Economic Review 101 (5), 1933–1963.
Galasso, A., Schankerman, M., 2013. Patents and cumulative innovation: causal evidence from the courts, CEP Discussion Paper No. CEPDP1205.
Gans, J., Murray, F., 2012. Funding scientific knowledge: selection, disclosure and the public-private portfolio. In: Lerner, J., Stern, S. (Eds.), The Rate and Direction of Inventive Activity Revisited. University of Chicago Press, Chicago, IL, Chapter 1.
Gavetti, G., Levinthal, D.A., Rivkin, J.W., 2005. Strategy making in novel and complex worlds: The power of analogy. Strategic Management Journal 26 (8), 691–712.
Green, J., Scotchmer, S., 1995. On the division of profit in sequential innovation. The RAND Journal of Economics 26, 20–33.
Gowers, T., Nielsen, M., 2009. Massively collaborative mathematics. Nature 461, 879–881.
Hart, O., 1995. Firms, Contracts, and Financial Structure. Oxford University Press.
Haeussler, C., Jiang, L., Thursby, J., Thursby, M., 2009. Specific and general information sharing among academic scientists, NBER Working Paper Series 15315.
Heller, M.A., Eisenberg, R.S., 1998. Can patents deter innovation? The anticommons in biomedical research. Science 280 (5364), 698–701.
Henkel, J., Baldwin, C., 2010. Modularity for value appropriation: how to draw the boundaries of intellectual property. Harvard Business School Working Paper.
Huang, K.G., Murray, F.E., 2009. Does patent strategy shape the long-run supply of public knowledge? Evidence from human genetics. Academy of Management Journal 52 (6), 1193–1221.
Kauffmann, S.A., 1993. The Origins of Order: Self-organization and Selection in Evolution. Oxford University Press, Oxford, UK.
Katz, A., 2014. The First Sale Doctrine and the Economics of Post-Sale Restraints. Brigham Young University Law Review 55. (1).
Khatib, Firas, et al., 2011. Algorithm discovery by protein folding game players. Proceedings of the National Academy of Sciences 108 (47), 18949–18953.
Kitch Edmund, W., 1977. The nature and function of the patent system. The Journal of Law and Economics (October) 20, 265–290.
Kuhn, T., 1962. The Structure of Scientific Revolutions. University of Chicago Press, Chicago, IL.
Lakhani, K., Boudreau, K., Loh, P.-R., Backstrom, L., Baldwin, C., Lonstein, E., Lydon, M., MacCormack, A., Arnaout, R., Guinan, E., 2013. Prize-based experiments can provide solutions to computational biology problems. Nature Biotechnology 31, 108–111.
Lakhani, K., Wolf, R., 2005. Why hackers do what they do: Understanding motivation and effort in free/open source software projects. In: Feller, H., Fitzgerald, B., Hissam, S.A., Lakhani, K.R. (Eds.), Perspectives on Free and Open Source Software. MIT Press, Cambridge, MA.
Lakhani, K., von Hippel, E., 2003. How open source software works: "free" user-to-user assistance. Research Policy 32 (6), 923–943.
Laursen, K., Salter, A., 2006. Open for innovation: The role of openness in explaining innovation performance among UK manufacturing firms. Strategic Management Journal 27 (2), 131–150.
Lerner, J., Tirole, J., 2002. Some simple economics of open source. The Journal of Industrial Economics 50 (2), 197.
Lerner, J., 2006. 150 years of patent office practice: law and institutions. American Law and Economics Review 7, 112–143.
Lessig, L., 1999. Code and Other Laws of Cyberspace. Basic Books, New York, NY.
Levinthal, D.A., 1997. Adaptation on rugged landscapes. Management Science 43, 934–950.
Maas, H., Wagenmakers, E.-J., 2005. A psychometric analysis of chess expertise. The American Journal of Psychology 118, 29–60.
March, J.G., 1991. Exploration and exploitation in organizational learning. Organization Science 2 (1), 71–87.
Marx, M., Strumsky, D., Fleming, L., 2009. Mobility, skills and the Michigan noncompete experiment. Management Science 55 (5), 875–899.
Mazzoleni, R., Nelson, R., 1998. The benefits and costs of strong patent protection: a contribution to the current debate. Research Policy 27 (3), 273–284.
Meyer, P.B., 2013. The airplane as an open source invention. Revue Economique 64 (1), 115–132.
Moldovanu, B., Sela, A., 2001. The optimal allocation of prizes in contests. American Economic Review 91, 542–558.
Moser, P., Nicholas, T., 2013. Prizes, publicity and patents: Non-monetary awards as a mechanism to encourage innovation. The Journal of Industrial Economics 61 (3), 763–788.
Moon, S., 2011. 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, 1–32.
Mukherjee, A., Stern, S., 2009. Disclosure or secrecy? The dynamics of open science. International Journal of Industrial Innovation 27, 459–462.
Murray, F., O'Mahony, S., 2007. Exploring the foundations of cumulative innovation: implications for organization science. Organization Science 18 (6), 1006–1021.
Murray, F., Stern, S., 2007. 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, 648–687.
Murray, F., Aghion, P., Dewatripont, M.M., Kolev, J., Stern, S., 2009. Of mice and academics: Examining the effect of openness on innovation, NBER Working Paper Series 14819.
Nordhaus, W.D., 1972. The optimal life of a patent. The American Economic Review 62 (3), 428–431.
Nelson, R., 1961. Uncertainty, learning, and the economics of parallel research and development. Review of Economics and Statistics 43, 351–368.
Nelson, R.R., Winter, S.G., 1982. An Evolutionary Theory of Economic Change. Harvard University Press, Cambridge, MA.
Newell, A., Simon, H.A., 1972. Human Problem Solving. Prentice-Hall, Engelwood Cliffs, NJ.
Nuvolari, A., 2004. Collective invention during the British Industrial Revolution: the case of the Cornish pumping engine. Cambridge Journal of Economics 28 (3), 347–363.
O'Mahony, S., 2003. Guarding the commons: how community managed software projects protect their work. Research Policy 32, 1179–1198.
O'Mahony, S., Lakhani, K.R., 2011. Organizations in the Shadow of Communities. In: Marquis, C., Lounsbury, M., Greenwood, R. (Eds.), Communities and Organizations, Vol. 33. Research in the Sociology of Organizations. Emerald Group Publishing, pp. 3–36.
Osterloh, M., Rota, S., 2007. Open source software development: just another case of collective invention? Research Policy 26 (2), 157–171.
Parker, G., Van Alstyne, M., 2010. Innovation, openness and platform control. In: EC'10 Proceedings of the 11th ACM Conference on Electronic Commerce.
Raymond, E., 1999. The Cathedral and the Bazaar. O'Reily and Associates, Sebastapol.
Roberts, J., Hann, I., Slaughter, S., 2006. Motivations, participation and performance in open source software development. Management Science 52, 984–999.
Rochet, J.C., Tirole, J., 2003. Platform competition in two-sided markets. Journal of the European Economic Association 1 (4), 990–1029.
Romer, P., 1990. Endogenous technological change. Journal of Political Economy 98 (5).
Rosenberg, N., 1982. Inside the Black Box: Technology and Economics. Cambridge University Press, Cambridge, UK.
Rosenkopf, L., Almeida, P., 2003. Overcoming local search through alliances and mobility. Management Science 49 (6), 751–766.
Rosenkopf, L., Nerkar, A., 2001. Beyond local search: Boundary-spanning, exploration and impact in the optical disk industry. Strategic Management Journal 22, 287–306.
Sahal, D., 1985. Technological guideposts and innovation avenues. Research Policy 14 (2), 61–82.
Salter, A.J., Martin, B.R., 2001. The economic benefits of publicly funded basic research: A critical review. Research policy 30 (3), 509–532.
Schankerman, M., 1998. How valuable is patent protection? Estimates by technology field. The RAND Journal of Economics 29 (1), 77–107.
Sampat, B., Williams, H., 2014. How do patents affect follow-on innovation? Evidence from the human genome, NBER Working Paper.
Scotchmer, S., 1991. Standing on the shoulders of giants: cumulative research and the patent law. Journal of Economic Perspectives 5 (1), 29–41.
Scotchmer, S., 2004. Innovation and Incentives. The MIT Press, Cambridge, MA.
Simon, H.A., 1962. The architecture of complexity. Proceedings of the American Philosophical Society 106 (6), 467–482.

Solow, R.M., 1957. Technical change and the aggregate production function. Review of Economics and Statistics 39 (3), 312–320.
Stephan, P., 1996. The economics of science. Journal of Economic Literature 34, 1199–1235.
Stephan, P., 2012. How Economics Shapes Science. Harvard University Press, Cambridge, MA.
Stern, S., 2004. Biological Resource Centers: Knowledge Hubs for the Life Sciences. The Brookings Institution Press, Washington, DC.
Taylor, C., 1995. Digging for golden carrots: An analysis of research tournaments. American Economic Review 85, 872–890.
Torrance, A.W., 2010. Synthesizing law for synthetic biology. Minnesota Journal of Law, Science & Technology 11, 629.
Utterback, J., Abernathy, W., 1975. A dynamic model of product and process innovation. Omega 3, 639–656.
Vincenti, W.G., 1994. The retractable airplane landing gear and the Northrop 'anomaly': variation-selection and the shaping of technology. Technology and Culture 35, 1–33.
von Hippel, E., 1988. The Sources of Innovation. MIT Press, Cambridge, MA.
von Hippel, E., von Krogh, G., 2003. Open source software and the "private-collective" innovation model: issues for organization science. Organization Science 14 (2), 209–223.
von Hippel, E., 2005. Democratizing Innovation. MIT Press, Cambridge, MA.
von Krogh, G., Spaeth, S., Lakhani, K.R., 2003. Community, joining, and specialization in open source software innovation: a case study. Research Policy 32 (7), 1217–1241.
Weitzman, M., 1998. Recombinant Growth. Quarterly Journal of Economics.
West, J., 2003. How open is open enough? Melding proprietary and open source platform strategies. Research Policy 32 (7), 1259–1285.
West, J., Lakhani, K.R., 2008. Getting clear about communities in open innovation. Industry and Innovation 15 (2), 223–231.
Williams, H., 2013. Intellectual property rights and innovation: evidence from the human genome. Journal of Political Economy 121, 127.
Zhang, X., Zhu, F., 2011. Group size and incentives to contribute: a natural experiment at Chinese Wikipedia. The American Economic Review 101 (4), 1601–1615.