International Journal of Industrial Organization 27 (2009) 449–462

Contents lists available at ScienceDirect

International Journal of Industrial Organization

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

Disclosure or secrecy? The dynamics of Open Science ☆

Arijit Mukherjee [a], Scott Stern [b,*]

[a] Bates White, LLC and Michigan State University, United States
[b] Kellogg Graduate School of Management, Northwestern University, United States

ARTICLE INFO

Article history:
Received 8 November 2007
Received in revised form 19 November 2008
Accepted 30 November 2008
Available online 6 December 2008

JEL classification:
O31
O32
L33

Keywords:
Open Science
Pasteur's Quadrant
Overlapping generations

ABSTRACT

Open Science is a dynamic system of knowledge production that depends on the disclosure of knowledge by researchers as an input into knowledge production by future researchers. To analyze the conditions supporting Open Science, we develop an overlapping generations model that focuses on the trade-off between disclosure and secrecy. While secrecy yields private returns that are independent of the actions of future generations, the benefits of disclosure depend in part on the use of disclosed knowledge by the subsequent researchers. We show that Open Science and Secrecy are both potential equilibria, and that the feasibility of Open Science depends on factors such as the costs of accessing knowledge from prior generations and the relative benefits to private exploitation under secrecy versus disclosure. In parameter regions where both Open Science and Secrecy can be supported, Open Science is associated with a higher level of social welfare. The analysis has policy implications for a number of areas, including public support for research training, appropriate design of formal intellectual property, and the role of scientific norms and institutions (such as an effective peer review process) in maintaining Open Science over the long run.

1. Introduction

At least since the development of scientific societies and related research institutions in the seventeenth century, the centrality of cumulativeness in scientific and technical advance has been recognized, most famously by Newton, who observed that scientific progress depends on "standing on the shoulders of giants." While economic theory has focused on deriving the implications of cumulativeness for related economic variables such as the equilibrium growth rate (Romer, 1990; Grossman and Helpman, 1991; Jones, 1995; Jones, forthcoming) or the incentives for commercial innovation (Scotchmer, 1991; Gallini and Scotchmer, 2002; Scotchmer, 2004), relatively little research has focused on the microeconomic conditions that support a cumulative research environment.

The fact that knowledge is produced does not guarantee that follow-on researchers will be able to exploit that knowledge (Polanyi, 1967). Effective diffusion of knowledge across researchers and over time requires that individuals are aware of the extant knowledge and that they pay the costs of accessing that knowledge. The ability of a society to stand on the shoulders of giants depends not only the amount of knowledge it generates but also on the quality of mechanisms for storing knowledge, the trustworthiness of that knowledge, and the cost to future generations of accessing that knowledge (Mokyr, 2002; Furman and Stern, 2008).

Open Science is perhaps the most well-known system for achieving these objectives. Open Science is characterized by a distinctive set of incentives for cumulative knowledge production, including norms that facilitate disclosure and knowledge diffusion (Merton, 1973; Dasgupta and David, 1994). This system includes the recognition of scientific priority by future scientific generations, the importance of demonstrating experimental replicability, and a system of public (or coordinated) expenditures to reward those who contribute to cumulative knowledge production over the long term. By conditioning career rewards (such as tenure) on disclosure through publication, Open Science promotes cumulative discovery.[1] However, the logic underlying Open Science as an economic institution is more subtle. The ability to sustain disclosure over time depends not simply on the willingness of scientists to invest in research per se but also in their willingness to (1) invest in drawing upon the knowledge

☆ We thank Joshua Gans, Maria Goltsman, Tapas Kundu, Fiona Murray, Marcin Peski, Luis Vasconcelos, Michael Whinston, the Editor, two anonymous referees, and especially Julien Jamison for extremely helpful comments and suggestions. The first author gratefully acknowledges financial support from Bates White Research Enhancement Grant. The second author gratefully acknowledges funding from NSF Science of Science Policy Grant 0738394. All errors that may remain are ours.

* Corresponding author.
E-mail addresses: arijit@msu.edu (A. Mukherjee), s-stern2@northwestern.edu (S. Stern).

[1] Indeed, the policy rationale for public support of Open Science has been rooted in the idea that basic research is a public good and that ensuring an appropriate level of basic research requires some form of subsidy, most likely provided by the public sector (Nelson, 1959; Arrow, 1962). David (2004) builds upon and then extends a rich literature in the history of Science to emphasize that Open Science has long relied on the politically motivated patronage of key individuals. It is only within the past century that national governments have taken the lead in providing stable and extensive funding for Open Science (Nelson and Rosenberg, 1994; Mowery and Rosenberg, 1998).

0167-7187/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.ijindorg.2008.11.005

provided by prior researchers and (2) disclose their own discoveries in a way that can be accessed and exploited by future researchers.[2]

The ability to maintain Open Science may be challenged when discoveries are not only of scientific interest but also have significant commercial application. When a single discovery has *dual* applications – it can serve as an input to future scientific research and be exploited directly for commercial gain – a trade-off arises between the incentives to disclose through the scientific literature and the incentives to maximize direct commercial exploitation (Rosenberg, 1990; Stokes, 1997). Consider the Oncomouse (Murray, 2006). In the early 1980s, Professor Phil Leder at the Harvard Medical School developed the first genetically engineered mouse; it was called the Oncomouse. Leder and his colleague had used newly emerging transgenic techniques to insert an oncogene into a mouse embryo; the result was a mouse that was highly susceptible to cancer. Using the mice to examine the importance of oncogenes in the onset of cancer, Leder came to recognize that "it could serve a variety of different purposes, some purely scientific others highly practical" (Kevles, 2002, p. 83). This research was published in *Cell* in 1984, and, in 1988, a broad patent for the Oncomouse was granted by the USPTO. Harvard's licensee DuPont aggressively enforced these rights, including demands for "reach-through" rights and review of publications that used the Oncomouse in further scientific research. Over the next decade, a number of controversies surrounded the access to and credit for discoveries based on the Oncomouse. The conflict over the Oncomouse centered on the ability of the broader scientific community to exploit the Oncomouse (and to provide informal recognition to Leder and his coauthors) versus the incentives of DuPont to limit the diffusion of the Oncomouse in order to maximize its commercial advantage (Murray, 2006).

Although traditional models of science and innovation have often assumed a sharp delineation between purely scientific research and commercial applications, qualitative studies of scientific research have increasingly emphasized the importance of dual-use research (Rosenberg, 1974; Stokes, 1997; Murray, 2002; Murray and Stern, 2007). Stokes, in particular, suggested that a significant share of all scientific research combines the scientific and commercial motives and results in knowledge production in "Pasteur's Quadrant."[3] Pasteur's fundamental insights into microbiology simultaneously had practical applications for cholera and rabies while also serving as the foundation for the germ theory of disease (Geison 1995; Stokes, 1997).

This paper analyzes the feasibility of Open Science when research is conducted in Pasteur's Quadrant (i.e., has both scientific and commercial importance). We consider how incentives for access to prior knowledge, investment in knowledge, and the disclosure of discoveries depend on the disclosure and investment decisions of prior researchers and the access decisions of future researchers. A critical ingredient of our analysis is the fact that the incentives of any one researcher to participate in Open Science depend crucially on the choices of other researchers — i.e., the incentives to publish research in an academic journal depend on future researchers building on that discovery and providing appropriate citations to it in their own research. We model scientific disclosure as an endogenous economic outcome of the microeconomic environment, with the potential for Open Science depending on strategic interaction among researchers in their access, investment, and disclosure decisions.

Our model highlights two features of Open Science: (1) the ability to draw upon prior (disclosed) research and (2) the fact that the incentives to produce and disclose abstract knowledge depend on receiving credit from follow-on researchers. In contrast, the incentives for commercially motivated knowledge production are premised on the ability to limit the use of knowledge by others; we call this approach "Secrecy." Of course, the private returns to scientific research crucially depend on several exogenous factors such as the institutional and legal environment of the time. In our model, this is achieved through trade secrecy. (We consider the role of formal intellectual property rights (IPR) in an extension.) We embed the choice between secrecy versus disclosure into an overlapping generations framework in which each generation is composed of a single researcher who lives for two periods. During his first period of life, each researcher produces a knowledge output by choosing (1) whether to draw upon knowledge (if available) produced by the previous generation, (2) the level of investment in his own research, and (3) whether to disclose the produced knowledge for follow-on researchers in the next period. Each researcher faces a fixed cost of drawing upon prior knowledge, and a constant marginal cost of investment in his own research. The benefits to each researcher are composed of (1) the benefits from citations to his research by the next generation (if he chooses to disclose, and the next generation chooses to build on that research) and (2) private rents from proprietary exploitation of his knowledge. Researchers face a trade-off between maximizing the benefits from *private* exploitation (through secrecy) and earning a lower benefit from private exploitation but earning additional benefits from disclosure through the institutions of Open Science.

We draw out the equilibrium implications of this choice between secrecy and disclosure and focus on three potential outcomes: (1) "Open Science," in which each generation invests in access to prior knowledge, chooses a constant level of investment, and discloses knowledge to the next generation; (2) "Secrecy," in which each generation does *not* build on the knowledge produced by the prior generation, chooses a constant level of investment, and chooses not to disclose the knowledge produced to the subsequent generation; and (3) k-period "cycle" equilibrium, in which a single period of "Secrecy" is followed by $k-1$ periods of "Open Science."

At least one of these three types of equilibria *must* exist for any set of parameter values that describes the microeconomic environment. With that said, the feasibility of a given equilibrium depends crucially on the parameters of the economic environment. For example, the viability of Open Science is decreasing in both the cost of accessing knowledge produced by prior generations and in the relative benefits to private exploitation under secrecy versus disclosure. We also examine the role of factors such as the effectiveness of scientific institutions in promoting the effective transfer of knowledge across generations and the marginal cost of research investment. Rather than being grounded in differences in the type of knowledge produced, the model suggests that the feasibility of Open Science depends on the institutional and microeconomic environment in which that knowledge is produced; these parameters are themselves functions of the policy environment.

The model also highlights the potential for multiple equilibria for a given set of parameters, so that the choice between "Open Science" and "Secrecy" is endogenous to the strategic interaction among researchers. When multiple equilibria exist, we are able to rank welfare. Open Science, whenever viable, generates more surplus than any regime involving Secrecy. Moreover, among the set of Open Science equilibria, welfare increases as a function of the level of research investment. Finally, we considers a number of extensions and implications of the model: (1) the potential for knowledge spillovers across multiple generations (relaxing our assumption in the baseline model that spillovers only occur across immediately adjacent research generations), (2) the potential for hysteresis (is it more difficult to establish Open Science as an equilibrium than to maintain that equilibrium once it is established?), and (3) the role of formal intellectual property rights such as patents. The contribution of this paper is to isolate the equilibrium implications of the trade-off that arises for each research generation between secrecy and disclosure

[2] As has been emphasized by, among others, Blumenthal et al (1997), scientific researchers often withhold key materials or tools from follow-on researchers. This results in increasing policy concerns over access and transparency in the scientific commons.

[3] As in contrast with the knowledge produced for fundamental scientific interest (referred to as "Bohr's Quadrant") and the knowledge produced primarily for commercial gain (referred to as "Edison's Quadrant").