ARTICLE IN PRESS

Research Policy xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Research Policy

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

Above a swamp: A theory of high-quality scientific production[☆]

Bralind Kiri[a], Nicola Lacetera[b,*], Lorenzo Zirulia[c]

[a] University of Extremadura, Spain
[b] University of Toronto (Institute for Management and Innovation), Canada
[c] University of Bologna, ICRIOS – Bocconi, and Rimini Centre for Economic Analysis, Italy

ARTICLE INFO

Keywords:
Economics of science
Reliability of research
Research incentives
Science policy

ABSTRACT

We elaborate a model of the incentives of scientists to perform activities of control and criticism when these activities, just like the production of novel findings, are costly, and we study the strategic interaction between these incentives. We then use the model to assess policies meant to enhance the reliability of scientific knowledge. We show that a certain fraction of low-quality science characterizes all the equilibria in the basic model. In fact, the absence of detected low-quality research can be interpreted as the lack of verification activities and thus as a potential limitation to the reliability of a field. Incentivizing incremental research and verification activities improves the expected quality of research; this effect, however, is contrasted by the incentives to free ride on performing verification if many scientists are involved, and may discourage scientists to undertake new research in the first place. Finally, softening incentives to publish does not enhance quality, although it increases the fraction of detected low-quality papers. We also advance empirical predictions and discuss the insights for firms and investors as they "scout" the scientific landscape.

> Science does not rest upon solid bedrock. The bold structure of its theories rises, as it were, above a swamp. It is like a building erected on piles. The piles are driven down from above into the swamp, but not down to any natural or "given" base; and when we cease our attempts to drive our piles into a deeper layer, it is not because we have reached firm ground. We simply stop when we are satisfied that they are firm enough to carry the structure, at least for the time being.
>
> Karl R. Popper, The Logic of Scientific Discovery (1959, p. 111).

1. Introduction

The production of reliable and high-quality scientific research is valuable not only within the ivory tower of academia. Firms and investors, for example, assess opportunities also on the basis of the science underlying a new business idea, and "scout" the scientific landscape in search for discoveries that are scientifically sound and commercially promising (Baum and Silverman, 2004; Merck, 2015; Pfizer, 2015; Ryan, 2013). More broadly, scientific knowledge is a powerful engine of economic growth and social welfare (Romer, 1990; Stephan, 2012).

For these reasons, the debate about the reliability of research involves not only the scientific community, but also firms, policymakers and the public opinion. Several accounts have pointed to a "reproducibility crisis" in science (Allison et al., 2016; The Economist, 2013). In psychology, for example, a project attempting to replicate 100 studies succeeded only in 39 cases (Open Science Collaboration, 2015). Begley and Ellis (2012) reported that they could replicate only 6 out of 53 studies in oncology and haematology, and in a meta-analysis of genetic associations studies, Ioannidis et al. (2001) found that the results of the first study, often suggesting a stronger genetic effect, correlated only weakly with subsequent research. The social and economic costs of this lack of reliability may be substantial; Freedman et al. (2015), for example, estimate that every year 28 billion dollars are spent in the US on preclinical research that is not reproducible.

Science may "go wrong" for outright fraud or mistakes that, if major and detected, lead to retraction from publication (Azoulay et al., 2015a,b; Broad and Wade, 1982; Furman et al., 2012; Lacetera and Zirulia, 2011; Lu et al., 2013). Incentives prevailing in scientific communities, such as the "publish or perish" imperative (Abelson, 1990; Giles, 2007), are often blamed for inducing to frauds or grave inaccuracies.

☆ We are grateful to the participants to several seminars (University of Bologna, University of Florence, TOBB University, University of Salamanca) and conferences (2014 Izmir Workshop on Scientific Misconduct and Research Ethics in Economics, 2014 EARIE Conference, 2014 BRICK Conference, 2014 DRUID Conference, 2015 Barcelona GSE Summer Forum, 2015 Italian Economic Association Conference, 2015 REER Conference, and 2016 Annual Meeting of the French Economic Association) for their feedback. Astrid Marinoni provided outstanding research assistance.

* Corresponding author.
E-mail addresses: bkiri@unex.es (B. Kiri), nicola.lacetera@utoronto.ca (N. Lacetera), lorenzo.zirulia@unibo.it (L. Zirulia).

https://doi.org/10.1016/j.respol.2018.01.011
Received 15 May 2017; Received in revised form 14 December 2017; Accepted 13 January 2018

Please cite this article as: Kiri, B., Research Policy (2018), https://doi.org/10.1016/j.respol.2018.01.011

In a less pessimistic view, however, flaws, limitations and mistakes in a study just occur as "natural" steps toward better theories and findings (Aschwanden, 2015). Karl Popper's view of science, for example, holds that a finding or theory can be defined as scientific to the extent that it is falsifiable (Popper, 1959). Therefore, at each given time, the body of scientific knowledge includes findings that are limited or flawed in some ways, with corrections and improvements occurring as long as new results, confirming or falsifying the original ones, accumulate (Howson and Urbach, 1989). Building upon previous research and potentially identifying its limitations thus appears essential for a healthy working of the scientific community (Carpi and Egger, 2011).

The history of science provides many examples of how subsequent research challenged accepted findings. In some cases, improvements and corrections (or sometimes full-blown controversies) led to a better understanding of a phenomenon. For instance, the Copernican revolution benefited from critiques to some of its aspects, even if those critiques were based on wrong theories, such as Tycho Brahe's observations about inconsistencies in the heliocentric view (Sherwood, 2011). In other cases, such as the research on HIV and AIDS, advances occurred through progressive criticisms and falsifications of earlier results obtained with less reliable empirical strategies (Holmberg, 2008). In climatology, there is increasing agreement about the anthropogenic nature of climate change. However, counterarguments and evidence of scholars who are more skeptic are contributing to improve the overall reliability of research in this area (Sherwood, 2011). Critical views may be particularly valuable when scientific results attract media attention, as was recently the case in paleoanthropology following the discovery of Homo Naledi (Lents, 2016).

In other instances, research that built on previous work led to discarding that earlier work entirely; examples include polywater and cold nuclear fusion (Rousseau and Porto, 1970; Taubes, 1993). Livio (2014) describes "blunders" by some great scientists. Darwin's theory of evolution, for example, presented in its initial versions some flaws that Fleeming Jenkin, a Scottish engineer, pointed out, with his critique containing, in turn, some limitations subsequently reported by Arthur Sladen Davis; the contributions of Linus Pauling, the Nobel Laureate for Chemistry in 1954 (and for Peace in 1962), to the definition of the DNA structure were soon identified as flawed by Crick and Watson. Catalini et al. (2015) find that articles in immunology receiving "negative" citations (i.e. citations that criticize or limit the validity of a study) tend to be highly cited overall and therefore more prominent and relevant; in turn, papers making negative citations are not marginal (again as measured by overall citations).[1]

Based on these premises, this paper proposes a game-theoretic analysis of the interplay between the incentives to exert scientific effort and provide accurate results on the one hand, and the incentives to verify the validity of previous findings on the other hand. With our model, we address the following questions:

- What are the incentives of scientists to perform research on existing, established topics and to potentially exert control and criticism?
- Will these activities always improve upon or correct previous findings, or shall we expect some degree of imprecision at any given time?
- How do these incentives interact with those to produce novel, high-quality findings?
- What factors determine the incidence of imperfect science?
- Which policy interventions could improve the reliability of science? Which policies, instead, would be ineffective or even counter-productive?

The main result from the basic version of the model, which we describe in Section 2 and solve in Section 3, is that a certain fraction of low-quality scientific knowledge characterizes all the equilibria of the game. Incentives to verify findings may be too low, thus reducing also incentives to perform high effort to produce reliable research; or they may be high enough to lead to verification with positive probability, and in turn, to the production of higher-quality research on average. An implication of this result is that never observing low-quality research in a scientific field may be due to a lack of verification activities and, as such, can be a source of concern rather than a signal of the solidity of a body of knowledge. Therefore, fields that display controversies may indicate greater health and promise than fields without these debates.

Although our result suggests that observing a certain fraction of flawed science may indeed be a natural and desirable feature, the identification of those characteristics associated to higher reliability allows to assess different policies meant to increase the overall reliability of science, as well as the ability to sort scientific results of different quality. We do this in Section 4, where we perform comparative statics exercises on the basic version of the model and we extend it in several directions.

We show that reducing the value of a publication for the knowledge originator does not have an impact on research quality, although it increases the fraction of low-quality papers that are identified. Conversely, reducing the costs (or increasing the benefits) for scientists to verify the results of others increases the overall expected quality of research. These findings highlight an important role for incremental research that reinforces, limits, or even just confirms previous results; they also suggest that a softening of publish-or-perish attitude would simultaneously increase research quality and the identification of low-quality research if acting at the same time on the incentives to produce and verify knowledge.

We also identify, however, a few countervailing effects of enhancing verification activities. For example, less costly (or too frequent) verification activities may lead a scientist to not undertake a new, potentially socially valuable research project in the first place. Thus some level of "protection" of one's research (e.g. concerning policies for data sharing) might be desirable in certain cases.

Additional results concern the impact of the size of the scientific community. The performance of verification activities by a high number of scientists may lead to the overall reduction of these activities and of the expected quality of research if individual rewards from scrutiny are lower because they are shared among colleagues. Also, in scientific communities where interactions are repeated and frequent, scientists may "collude," i.e. avoid to verify each other's research.

A final set of policies that we consider in the model regard the direct involvement of journals in certifying the reliability of research. We derive that this involvement may reduce reliability by crowding out the scientists' incentives to perform verification activities; in contrast, attention of journals to other aspects of quality would generally improve reliability.

In Section 5 we outline insights for companies and investors interested in exploring the scientific landscape for business opportunities, we propose a few empirical tests based on the model's prediction, and we extend the application of the framework to other contexts beyond the working of the scientific community.

Related literature. Two early contributions that analyze replication activities formally are Mirowski and Sklivas (1991) and Wible (1998). Mirowski and Skivas model the interaction between an originator of knowledge and a potential replication, plus a set of potential extenders. In their analysis, (exact) replication never occurs unless editors require the originator to reveal a high enough level of information about their work, whereas extensions are more likely to occur in equilibrium. Wible proposes an application of Becker's consumption-production theory to the time allocation of a scientist into genuinely replicable articles and seemingly replicable articles, the

[1] A more limited analysis that we conducted on 1037 articles on climate change published in *Nature* (between 1975 and early 2015) and *Nature Climate Change* (2007–early 2015) shows about 215 cases in which some papers were negatively cited.