THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITE DE TOULOUSE

Délivré par l'Université Toulouse Capitole

École doctorale : **Toulouse Sciences Economiques**

Présentée et soutenue par

BERNARD Olga

le 7 décembre 2020

Essais en économie de la science

Qualité de publication et incitations à faire de la recherche

---

Discipline : **Economie**

Spécialité : **Sciences Economiques**

Unité de recherche : **UMR 5314 -TSE-R**

Thèse dirigée par : Catherine Bobtcheff et Thomas Mariotti

JURY

Mme Lucie MENAGER, Rapporteure
M. Nicolas CARAYOL, Rapporteur
M. Yassine LEFOUILI, Examinateur
M. Alexandre DE CORNIERE, Examinateur
Mme Catherine BOBTCHEFF, Directrice de thèse
M. Thomas MARIOTTI, Co-directeur de thèse

Essays in Economics of Science

OLGA BERNARD

Composition du Jury :

Lucie Ménager Professeure, Université Paris II	Rapporteur
Nicolas Carayol Professeur, Université de Bordeaux	Rapporteur
Yassine Lefouili Professeur, Ecole d'Economie de Toulouse	Examinateur
Alexandre de Cornière Professeur, Ecole d'Economie de Toulouse	Examinateur
Catherine Bobtcheff Professeure, Ecole d'Economie de Paris	Directrice de thèse
Thomas Mariotti Professeur, Ecole d'Economie de Toulouse	Co-directeur de thèse

Acknowledgments

I am very grateful to Catherine Bobtcheff and Thomas Mariotti, who supervised my thesis. They taught me to be well organized and kept pushing me forward until the end. Catherine is very demanding and at the same time very involved in my thesis. She was always here when I needed and strongly supported me during all these five years. Thomas has also high research standards and a very good intuition, which helps me a lot for the research direction to take and explain the results. The mix of these two supervisors enables me to reach a high level of requirements and give my best until the end.

I also thank Lucie Ménager, Nicolas Carayol, Yassine Lefouili and Alexandre de Cornière, Catherine Bobtcheff and Thomas Mariotti, who made me the honor to be part of the Jury of this thesis.

Moreover, I give a special mention for my husband Jean-Marc and my daugther Jeanne for their forceful mental support. Jean-Marc is a wonderful husband who was always here in the good as well as in the bad moments. His support and strength are way out of the norm. My daugther is so precious that she gave me all the energy which I needed to combine both my role as a mother, without any help apart my family and husband, and my thesis, who appears to be my second child (laugh)!

Besides, I warmly thank my parents, sister and family-in-law for having supported me, especially since my daughter was born. I is not easy as a mother to manage both work and family duties but they helped me reaching a high level of requirements for these two wonderful tasks.

In addition, I thank all the people regularly attending the Applied Theory Workshop at Toulouse School of Economics, especially Takuro Yamashita, Daniel Garrett, Johannes Hörner, François Salanié, Bertrand Gobillard, Catherine Bobtcheff, Thomas Mariotti, Roberto Sarkisian, Niccolo Lomys, Christopher Sandmann and Andres Salamanca. Thanks to them, I extended my research knowledge and they gave me useful advices for my research work.

Finally, I thank all my colleagues met during the PhD, especially Yuting Yang, Jacopo Bregolin, Christopher Sandmann, Ling Zhou Joana Duran, Julia Hoefer and Hussein Bidawi. Life as a PhD was happier and evenings longer with them. The list would be too long if I had to cite everyone!

Contents

Introduction

1 Why studying the economics of science?

Special attention has to be given to science by economists. Although the implementation of research results takes time to reach the common use, the latter has a direct impact on economic growth. For instance, medical research has enabled to expand considerably life expectancy, as underlined by The Antiretroviral Therapy Cohort Collaboration (2008) for antiretroviral therapies. Furthermore, as studied by Rosenberg and Mokyr, the relationship between science and economic growth is complex and non-linear. The first reason is that, as equipment plays a major role in scientific discovery, economic growth has an impact of the speed at which science is produced. The second reason is that technological breakthroughs play a key role in encouraging scientists in the public sector to develop new programs and research agendas. In addition, an interesting aspect of science is its social value.

1.1 The social value of science

"If nature has made any one thing less susceptible than all others of exclusive property, it is the action of the thinking power called an idea, which an individual may exclusively possess as long as he keeps it to himself; but the moment it is divulged, it forces itself into the possession of every one, and the receiver cannot dispossess himself of it. Its

peculiar character, too, is that no one possesses the less, because every other possesses the whole of it."

In this citation, Arrow explains in 1990 (1967 edition) why knowledge shares the properties of a public good, which was first noticed by Jefferson. Knowledge is not depleted when shared, and once it is made public others cannot easily be excluded from its use. Besides, adding one user is almost costless. Interestingly, and unlike other public goods, the stock of knowledge increases when the number of users increases. This implies that the transmission of knowledge is a positive sum game, as denoted by Foray (2004).

However, in the competitive research market, the non-excludability nature of knowledge allows free-riders to join the game. This provides too low incentives for scientists. Nevertheless, the current science reward system counterbalances the lack of incentives created by the public nature of knowledge.

1.2 Understanding the scientist's incentives

To begin with, a major incentive for scientists is that they aim at establishing the priority of a discovery. Merton (1957) was the first highlighting this fact, explaining that it has been a constant characteristic of science for more than three centuries. Rewarding priority can take several forms. For instance, eponymy attaches the name of the scientist to the discovery. Many scientists have left their mark: Planck's constant, the Copernician system, and more recently, the Higgs particle. In addition, rewards can take the form of prizes. The Nobel Prize is the most famous, but many others exist, such as the Shaw Prize and the Spinoza Prize, which represent more than a million US dollars each. Another and more common way of rewarding priority is publication. To that extent, the contribution of a scientist is often measured by using the number of citations of the latter's publications.

Nevertheless, the fight for priority implies that scientists are incentivized to disclose their work quickly. The delay between writing an article and submitting it can be very short, as well as the delay between the receipt of a manuscript and its publication. For example, the journal *Science* asks that referee reports should be returned within 7 days of receipt so that the latter publishes quickly following the editorial decision to accept. When two scientists disclose similar discoveries, the fight can be harsh to argue for priority. For instance, Newton took extreme measures to establish that he, not Leibniz, was the inventor of the calculus, as stated by Merton (1969). In this respect, science is often viewed as a “winner-take-all” contest, leaving only little rent for the followers. This scheme is partly due to the fact that scientific effort is not observable, as underlined by Dasgupta (1989). However, the followers can also add social values, by undertaking replication and verification activities. Taking the example of cancer cure, many treatments can be found, each of them bringing added value to society.

Furthermore, monetary rewards is another strong incentive to keep scientists innovating. These rewards split in two kinds. The first kind does not depend on the scientist’s success in the priority race. It compensates for the risk taken when searching for an uncertain result. The second kind is attributed to the winner and reflects her contribution to science. In this respect, counts of publication and citations play a major role in the allocation of academic promotions, as studied by Hicks (2007). As a matter of fact, scientists working in academia are often fully insured and earn roughly the same salary, no matter their place in the priority race. The winner position is rather rewarded by other monetary forms, such as a prize money, speaking and consulting fees.

Finally, another type of reward which boosts the scientist’s motivation is the satisfaction derived from solving the puzzle. In this respect, Feynman (1999), a 1965 Nobel Prize winner, said: “I don’t see that it makes any point that someone in the Swedish Academy decides that this work is noble enough to receive a prize. I’ve already got the prize. The prize is the pleasure of finding the thing out, the kick in the discovery..." This sort of reward makes science a worldsize game in which players are scientists. As underlined by Stephan (2010), future research should make a "special emphasis [...] on the public nature of knowledge and characteristics of the reward structure that encourage the production and sharing of knowledge".

2 The quality of publications

These last decades, the quality of published works has been placed on the top list of science agenda. Now more than ever, the latter has been questionned with the Covid-19 crisis.

2.1 Why is it a major issue?

To begin with, fake news can be spread very quickly. To that extent, new results should be considered carefully. Indeed, Ioannidis (2020) emphasizes that, based on Altimetric Scores, the most discussed and visible scientific article over all the articles publishes in the past eight years is a preprint from Pradhan et al. (2020), which asserts that the Covid-19 protein shares similarities with the HIV-1 protein. Even though the authors retracted the paper, the latter keeps making noise by stirring the passions. Furthermore, Kupferschmidt (2020) reports that the *New England Journal of Medicine* published last January a study claiming that a first asymptomatic case has transmitted the virus. Nevertheless, researchers did not ask, but the latter had symptoms.

Moreover, it seems that hot topics articles are more likely to be published faster than the others, especially those which concern transmission, morbidity and mortality. Palayew et al. (2020) alarm that the way Covid-19 articles were published "raises concerns about the quality of the evidence base and about the risk of misinformation being spread with harmful consequences". Focusing on published and preprint reports of prediction models for diagnosing of Covid-19, Wynants et al. (2020) show that models are poorly reported and the reported performance too optimistic, with a high risk of bias. They ask for more rigor and adherence to the TRIPOD (transparent reporting of a multivariable prediction model for individual prognosis or diagnosis) reporting guideline.

Besides, a large volume of research has been produced since the beginning of the pandemic crisis, in a surprisingly short amount of time. In this respect, Da Silva et al. (2020) report that over 23.500 articles were published between January and June 2020. As analyzed by Palayew et al. (2020), an average of 367 Covid-19 journal articles were published per week, from February to April 2020. Surprisingly, the median time from submission to acceptance was just six days. Comparing these results with the 2014 Ebola pandemia, only four articles per week were published from August to October 2014, with a mean time to acceptance of 15 days. According to Himmelstein (2016), the median delay from submission to acceptance is generally around 100 days. All these facts question the quality of the research published.

In addition, the publications quality issue is strongly related to the publication bias, which Brodeur et al. (2016) highlight. Using test statistics from studies published between 2005 and 2011 in the *American Economic Review*, the *Journal of Political Economy* and the *Quarterly Journal of Economics*, they show that the observed distribution of tests has a weird two-humped camel shape, with missing p-values between 0.25 and 0.10. They suggest that scientists inflate the value of just-rejected tests by choosing particular specifications. Furthermore, Ioannidis (2005a) underlines that "for many current scientific fields, claimed research findings may often be simply accurate measures of the prevailing bias".

Over the last decade, many attempts have been made to fasten the publication process, which is deplored to be long. The impact on the publication quality is unclear.

A first attempt is the pre-print option, which is a version of an article that the author submits to the editorial board of a scientific journal. The article is then reviewed by researchers who evaluate the content of the article and ask for clarifications. This stage is called peer-review. In this respect, journals like *eLife* and *The Journal of Clinical Investigation* announced that authors will only be asked for revisions to clarify and better present their results. In addition, *eLife* encourages the authors to post their submissions as preprints. This could be an answer to the current unavailability of reviewers who are focused with their own research and do not receive direct benefits from reviewing. Besides, journals offer a fast-track publication option, which enables scientists to pay a higher cost to get reviewed and published faster. Despite journals claim that the quality is as good as standard publication process, further investigation must be done.

Various solutions could be implemented to answer to the worrying publication quality. For instance, journals could enforce strong standards during public-health emergencies to avoid possibly poor research publication quality. In this respect, the European Association of Science Editors (EASE) "urges all involved in collecting and publishing data related to the pandemic to adhere to ethical guidelines, and to follow standard reporting guidelines, for example CONSORT for clinical trials and STROBE for epidemiological studies" (Bazdaric and Smart, 2020). Besides, effort should be made to develop multiple independent quality-assured systems, which relies more on public fundings than on the private sector. To this purpose, several curated databases were elaborated, such as LitCovid (US National Library of Medicine), Novel Coronavirus Research Compendium (Johns Hopkins Bloomberg School of Public Health) and Publons (Clarivate Analytics).

2.2 The replicability problem

Another way to increase the publications quality is to check the articles' outcomes after their publication. After a publication, if another article proves the same results with same or different methods and data, it is called a "replication". If instead, the results of the new study are not statistically significant or contradict those of the initial study, it is called a "refutation". As a matter of fact, Ioannidis (2005a) highlights that science is going through a "reproducibility crisis".

Indeed, many replication attempts have failed. One critical concern is public health care. In this respect, the publications quality has a direct impact on the way patients are treated, in other words, it is a matter of life or death. To that extent, Ioannidis (2005b) analyzes clinical studies of three major general clinical or specialty journals between 1990 and 2003. He finds that 44% of 45 medical studies are replicated. Looking at cancer drugs, Davis et al. (2017) prove that 57% of 68 drugs entered the European market without evidence of benefit on survival or quality of life. Besides, after a minimum of 3.3 years after their launch, there was still no conclusive evidence that these drugs have improved the quality or improved the patients' life. Using multiple random sets, Mischiels et al. (2005) study the stability of microarray gene-expressions to predict cancer outcome. They find that the list of genes identified as predictors of prognosis was highly unstable and that the outcomes depends strongly on the selection of patients in the training sets. All these studies mention a poor research replicability.

The replicability issue reaches every research field and the publications' results are often over-estimated. In biomedicine, Iqbal et al. (2016) study a random sample of 441 biomedical journal articles published between 2000 and 2014. As a matter of fact, there were only four replication studies and only 16 studies with data including in a systematic review or meta-analysis. [1] This shows both that there are not enough

[1] A meta-analysis is a statistical analysis that combines the results of multiple scientific studies. It

replication studies and a lack of transparency. In a meta-analysis performed on 370 genetic studies, Ioannidis et al. (2001) find statistically significant heterogeneity in 14 out of 36 groups of genetic association studies on the same topic with stronger effects in the first study of a topic than subsequent replication attempts in 25 cases.

Furthermore, the replicability rate varies accross disciplines. In cell biology, Begley and Ellis (2012) find that the latter is around 11%. In psychology, the Open Science Collaboration (2015), a conglomerate formed by scientists, find a rate of 50%. In economics, Camerer et al. (2016) study 18 experiment studies published in the *American Economic Review* and the *Quarterly Journal of Economics* between 2011 and 2014. They find that the replicability rate is between 67% and 78%, which is higher than in the other disciplines.

Finally, the publications outcomes directs the way money is spent in both the public and the private sectors. To that extent, reproducibility of research should be promoted. Sadly, Freedman et al. (2015) deplores that 28 billion dollars are spent each year in the US on preclinical research which is not reproducible. In this way, journals should incentivize replication and refutation. Some have encouraged transparency through codes of ethics, such as the Berkeley Initiative for Transparency in the Social Sciences (BITSS) and the Transparency and Openness Promotion (TOP).

In addition, several journals now devote part of their publications to replication studies, such as the *Royal Society Open Science*, the *Journal of Experimental Psychology: General* and the *Journal of Applied Econometrics*. Moreover, new journals have been fully dedicated to replication, such as ReScienceX, a free-to-publish, free-to-read and peer-reviewed journal recently created by *Nature*. In this respect, more attention is dedicated to replication and refutation.

occurs when there are several scientific studies dealing with the same issue, with each study reporting measurements which are expected to have some degree of error.

3 Science 2.0

Nowadays, the use of Internet is so common that we would not even think how to do without it. It has incredibly fasten the speed at which research is conducted and information is spread almost instantaneously all around the world. It was created only a few decades ago and yet has become necessary to undertake scientific research. To that extent, *Science 2.0* is a term used to describe the new way of making science with digitization. It is often related to *open science*, a movement which aims at making science available to everyone.

3.1 The new challenges of Science 2.0

With digitization come several concerns such as *open access*, *open data*, *transparency* and *collaboration*, as noticed by Dr. Stephane Berghmans (Kisjes, 2015). The first concern, open access, aims at providing worldwide access to printed and handwritten information. In this respect, many open-access journals have been created, which remove or reduce barriers to copying or reuse by applying an open license for copyright. For instance, the Public Library of Science (PLOS) is an open-access publisher which hosts several journals, such as PLOS Biology, created in 2003, and PLOS ONE, created in 2006. *Theoretical Economics* is a peer-reviewed open-access journal established in 2006.

The second concern, open data, should not be neglected. Data is a valuable product of science which is underutilized and shared, partly due to the stockage and hostage needs. The development of new tools to meet the needs implies considerable investment. Interestingly, digitization appears to be a valuable solution to make the cultural and historical heritage accessible to everyone and searchable through the Web. Once this heritage is digitized, it can be treated with scientific studies. In this respect, a digitization policy must be implemented, as underlined by Bountouri (2017). First, human and financial resources, as well as technological equipment, must be efficiently used to carry out archives' digitization. Second, a selection must be made beyound the huge amount of manuscripts. Third, laws must be enforced to frame the archives' digitization. Copyright is one of the many issues.

The third concern, transparency, is crucial in our highly interconnected world. As deplored by Lazer et al. (2018), its lack can generate fake news, which are unfortunately quickly spread worldwide. Furthermore, the transparency matter induces the creation of platforms to stock data. In this respect, Mendeley, a social network for scientists created in 2012, launched Mendeley Data, a platform in which researchers put their research data to be cited, shared and secured. In addition, journals, such as *MethodsX*, an open access journal, are created to enable scientist to publish their methods, thus facilitating their use by peers.

The fourth concern, collaboration, is inherent to the huge amount of data created and produces synergies. In this respect, the European Commission (EC) currently funds the Facilitate Open Science Training for European Research program (FOSTER), which was created in 2017 to help scientists using efficiently digital tools. In addition, the EC created the Open Science Policy Platform (OSPP) in 2016, which gives recommendations to foster and frame open science. One of them is to make research data Findable, Accessible, Interoperable and Reusable (FAIR). Furthermore, collaboration can invite non-experts to participate to research.

3.2 Citizen science

In this respect, citizen science (CS) has become a new and quite useful way of conducting research, which goes one step forward towards democratizing science. Also named "crowd science", "crowdsourced science" or "community science", CS requires the help of non-experts to implement scientific projects. Besides, the OSPP has made CS one of her top priorities in her research agenda. Interestingly, Theobald et al. (2015) estimate that 388 biodiversity CS projects contribute between $667 million to $2.5 billion annually.

Interestingly, some projects became very popular. This is partly due to the gamification of the CS project, which boosts the citizens' motivation. For instance, the FoldIt project was created in 2008 to help scientists studying proteins' shapes. Citizens contribute through an online game in which they use their spatial skills to design proteins. Two years after its launch, 200,000 citizens already participated. Eyewire, a CS project created in 2012 by a Princeton University scientist, is designed as a brain-mapping puzzle game to map retinal neurons. Over five years, 250,000 citizens from 150 countries have participated to that project.

Another famous project is Besides, platforms were developed to host CS projects. iNaturalist is a website jointly created by the California Academy of Sciences and National Geographic. Up to now, it has gathered more than 52 millions of observations about nature. The CS platform eBird, created in 2002, aims at collecting birds informations worldwide. Zooniverse is a platform created in 2007, which currently gathers 50 CS projects. A million citizens have already joined the website. This platform is actually an extension of the GalaxyZoo project, which aimed at classify galaxies by treating over a million of space data.

More recently, CS has been quite useful in the recent Covid-19 crisis. For instance, FoldIt launched last March a program to assist scientists to find a protein which will neutralize the virus. Besides, the platform EU-Citizen.Science proposes links to several citizen science projects which aim at find a cure to the virus. In addition, many online resources are available through open access and open science websites, including *The Lancet* and *Springer Nature*, which help the use of CS to fight against the pandemia.

Furthermore, Muki Haklay, a professor at the University College London, distinguishes different levels of citizens' participations. In the first one, called "crowdsourcing", citizens act as sensors, which means that they participate to environmental observation and data collection through the use of smartphones and networked devices. In the second one, called “distributed intelligence”, they act as basic interpreters. In the third one, called “participatory science”, they contribute to problem definition and data collection. In another level called “extreme citizen science”, citizens and scientists fully collaborate in problem definition, collection and data analysis.

However, there are some limitations which constrain the CS use. First, all projects are not always suited to CS if the research method is complex or if the work is very repetitive. In this way, some CS projects fail to be implemented because of limited participation. According to Cappa et al. (2018), this problem can be solved by giving to citizens correct incentives, such as monetary compensation and public online acknowledgment. In this respect, the way CS is conducted could follow a private-collective research model.

A second limitation is that the volonteer participation of non-experts may introduce a bias in the data, as noted by Lukyanenko et al. (2016). To answer this worrying CS projects’ quality, Kosmala et al. (2016) claim that successful projects lean on methods which increase data accuracy and account for bias. These methods include iterative project development, volunteer training and testing, expert validation, replication across volunteers, and statistical modeling of systematic error. Besides, Soroye et al. (2018) compared two datasets reporting butterflies species, one from a CS project called eButterfly, and the second coming from professionally collected observations. They show that the former is of high quality, which is due to the expert vetting process that eButterfly used. Besides, five new species were reported and geographic distribution information was improved for over 80% of species in the combined dataset when CS data was included. In addition, Gardiner et al. (2012), argue that the cost-effectiveness of citizen science data can outweigh data quality issues, if it is properly managed.

Furthermore, many impetus have been made to foster the use of CS. For instance, the Citizen Science Association, an organization which connects citizens to participate to scientific projects, created in 2016 an open-access journal called *Citizen Science: Theory and Practice*. Furthermore, CitizenScience.gov, an official government website, was created in 2016 to accelerate the use of crowdsourcing and CS across the U.S. government.

Finally, the impact of CS could go far beyond its first goal. In this respect, Bethany Brookshire, an American scientific journalist, writes “if citizens are going to live with the benefits or potential consequences of science [...], it’s incredibly important to make sure that they are not only well informed about changes and advances in science and technology, but that they also are able to influence the science policy decisions that could impact their lives”. To that extent, CS ethics should be considered, as it arises issues such as intellectual property. In this way, the European Citizen Science Association (ECSA) published in 2015 *Ten principles of Citizen Science*, which states that “the leaders of citizen science projects take into consideration legal and ethical issues surrounding copyright, intellectual property, data sharing agreements, confidentiality, attribution, and the environmental impact of any activities”.

4 Summary of chapters

4.1 Refutation in Research: How to Improve Publication Quality

In the first chapter of the thesis, [2] I examine how incentives for refutation affect publication quality. I build a sequential model of public experimentation with two scientists, a researcher (she) and a refuter (he). There exists a hypothesis which can be

[2]This chapter has been published in the *Annals of Economics and Statistics*.

either valid or invalid. The two scientists are uncertain about the hypothesis' validity and start the game with the same initial belief that the hypothesis is valid. At the beginning of the game, the researcher starts experimenting. She chooses when to publish a result confirming a hypothesis. While experimenting, she can observe for free some bad news which perfectly reveals that the hypothesis is invalid. If no bad news has arrived yet, the researcher's belief that the hypothesis is valid increases deterministically. The publication quality corresponds to the probability that the hypothesis is valid at the publication time. Once the result is published, the refuter starts experimenting but, unlike the researcher, he can choose between working on refutation or on an outside safe option at any time. His learning process is similar to the researcher. If he succeeds in refutation, the researcher incurs a refutation cost.

I find that the equilibrium belief at which the refuter stops experimenting and switches to the safe option is unique and does not depend on the initial prior. When scientists are initially very optimistic about the hypothesis' confirmation, none of the scientists experiments and the publication quality is equal to the initial belief. When they are initially pessimistic but the refutation cost is high, the researcher fully prevents the refuter from experimenting and the latter switches directly to the safe option after the result publication. When they are initially optimistic and the refutation cost is low, the researcher does not experiment but the refuter does: there is a refutation risk. When they are initially pessimistic and the refutation cost is high, the researcher experiments, but not fully deters the refuter from experimenting. There exists a refutation risk. In this case, the higher the refutation rewards, the lower the equilibrium publication quality in case the researcher is more efficient than the refuter. As a matter of fact, this result contradicts the literature's main policy recommendation, as Hamermesh (2007) stated. The opposite result holds when the refuter is more efficient. This second result is in line with Hamermesh (2007). When both scientists are equally efficient, refutation rewards have no impact on the publication quality.

In an extension, I study the case where the researcher's experimentation is private, which means that she can fraud and publish a result she knows to be wrong. Looking for a perfect Bayesian equibrium in pure strategies, the refuter cannot distinguish a researcher who received bad news from one who did not. I prove that the time during which the refuter experiments is fixed and does not depend on the publication time. In this respect, the researcher does not experiment and the publication quality is always equal to the intial belief. Hence, the publication quality is lower than in the public experimentation case, suggesting that transparency improves research quality. This last result is similar to Ozdenoren et al. (2019).

4.2 Citizen science vs. traditional science: the speed/quality trade-off

In the second chapter of the thesis, I examine the tradeoff of the scientist who has access to citizen science (CS) or traditional science (TS) to undertake her research. An infinite sequence of projects are successively available on a discrete time horizon. Each project is a public good with some value. There are three players, a scientist and two citizens. For each project, a scientist (she) chooses to implement it with CS or TS. With TS, she implements the project for sure but it takes one period. With CS, the project is implemented immediately, which is in line with Watson and Floridi (2016), but it requires the citizens' help. Each citizen (he) chooses to exert effort or not. The effort cost is privately observed by each citizen and depends on his motivation. If he is motivated, then effort is costless. If he is not motivated, then effort is costly. The implementation success of the project is increasing in the number of citizens exerting effort, as suggested by Nov et al. (2014).

In the unique citizens' Bayesian Nash equilibrium, a motivated citizen always exerts effort and a non-motivated one never does. The "CS quality" corresponds to the ex-ante probability that the project is implemented with CS. When the successive projects have the same value, the scientist's strategy is a cut-off: when the value of the project is above some threshold, she uses CS. Below this threshold, she uses TS. This result is generalized to any sequence of project values which satisfies the Markov property .

In an extension, two scientists compete to attract citizens on their project. The scientist's probability to attract citizens on her project increases in her investment cost and decreases in the cost invested by her competitor. I focus on stationnary symmetric Markovian equilibria, in which scientists make the same choice at every period. The scientist faces two types of choice to make. If her competitor has chosen TS for the last project, then she faces no competitor if she chooses CS for the current project. If her competitor has chosen CS for the last project, then they fight if both chooses CS for the current projects. As a matter of fact, I show that the equilibrium in pure strategies is unique: for sufficiently high project values, both scientists always choose CS. For sufficiently low values, both scientists always choose TS. In both cases, there exist no equilibrium in mixed strategies. For intermediate project values, the unique equilibrium is in mixed strategies: if one scientist is stuck with choosing TS at the former period, the other one always chooses CS. Otherwise, they both mix between TS and CS.

4.3 Citizen science and preemption

In the third chapter of the thesis, I study the tradeoff of a scientist choosing between citizen science (CS) and traditional science (TS) to undertake her research. After time 0, there exists a new idea which can be studied. Two scientists can discover this idea at any time. Their objective is to publish it before the other one does. However, competition is only potential as the time at which a scientist discovers the idea is not observable by her competitor. At the scientist's discovery time, she faces two technological choices, which are not available past that time. The first one is TS: she takes time to let the idea mature. There exists an optimal maturation delay which maximizes her publication payoff with TS without any competitor. With the second choice, called CS, the scientist involves the citizens' help to publish instantaneously the idea. However, she incurs a fixed cost to make the idea available to citizens, as reported by Bonney et al. (2009). Moreover, the latters are non-experts so there is some risk error that the publication is of bad quality, which brings her no payoff. This uncertain quality is asserted by Nov et al. (2014).

Focusing on Bayesian pure-strategy equilibria, I prove that there exist two kinds of symmetric stationary equilibria. When CS is low-cost, every scientist chooses CS. Otherwise, everyone chooses TS. Besides, I study equilibria in which every scientist chooses TS before a discovery time threshold and CS after. There exists no such equilibrium when it satisfies one of these two assumptions: i) the threshold discovery time is lower than the equilibrium publication time of the scientist discovering the idea at time 0 and ii) the equilibrium scientists' strategy is continuous at the threshold discovery time. At last, there exists no asymmetric equilibrium in which scientists choose different technological choices. These results suggest that the existence of the CS choice leaves no room for adaptation time. If CS is low-cost, even scientists discovering the idea early choose CS and have no time to adjust the technological change.

Bibliography

Bazdaric, Ksenija and Pippa Smart (2020) "ESE and EASE call for high standards of research and editing."

Begley, C. Glenn and Lee M. Ellis (2012) "Drug development: Raise standards for preclinical cancer research," *Nature*, 483 (7391), 531–533.

Bonney, Rick, Caren B Cooper, Janis Dickinson, Steve Kelling, Tina Phillips, Kenneth V Rosenberg, and Jennifer Shirk (2009) "Citizen science: a developing tool for expanding science knowledge and scientific literacy," *BioScience*, 59 (11), 977–984.

Bountouri, Lina (2017) *Archives in the digital age: Standards, policies and tools*: Chandos Publishing.

Brodeur, Abel, Mathias Lé, Marc Sangnier, and Yanos Zylberberg (2016) "Star wars: The empirics strike back," *American Economic Journal: Applied Economics*, 8 (1), 1–32.

Camerer, Colin F, Anna Dreber, Eskil Forsell, Teck-Hua Ho, Jürgen Huber, Magnus Johannesson, Michael Kirchler, Johan Almenberg, Adam Altmejd, Taizan Chan et al. (2016) "Evaluating replicability of laboratory experiments in economics," *Science*, 351 (6280), 1433–1436.

Cappa, Francesco, Jeffrey Laut, Maurizio Porfiri, and Luca Giustiniano (2018) "Bring them aboard: Rewarding participation in technology-mediated citizen science projects," *Computers in Human Behavior*, 89, 246–257.

Collaboration, Antiretroviral Therapy Cohort et al. (2008) "Life expectancy of individuals on combination antiretroviral therapy in high-income countries: a collaborative analysis of 14 cohort studies," *The Lancet*, 372 (9635), 293–299.

Dasgupta, Partha (1989) "The economics of parallel research," *The economics of missing markets, information, and games*, 129–48.

Davis, Courtney, Huseyin Naci, Evrim Gurpinar, Elita Poplavska, Ashlyn Pinto, and Ajay Aggarwal (2017) "Availability of evidence of benefits on overall survival and quality of life of cancer drugs approved by European Medicines Agency: retrospective cohort study of drug approvals 2009-13," *The British Medical Journal*, 359, j4530.

Feynman, Richard P (2005) *The pleasure of finding things out: The best short works of Richard P. Feynman*: Basic Books.

Foley, John P (1967) *The Jefferson Cyclopedia: A Comprehensive Collection of the Views of Thomas Jefferson Classified and Arranged in Alphabetical Order under Nine Thousand Titles Relating to Government, Politics, Law, Education, Political Economy, Finance, Science, Art, Literature, Religious Freedom, Morals, Etc*: Russell and Russell.

Foray, Dominique (2004) *Economics of knowledge*: MIT press.

Franzoni, Chiara and Henry Sauermann (2014) "Crowd science: The organization of scientific research in open collaborative projects," *Research policy*, 43 (1), 1–20.

Gardiner, Mary M, Leslie L Allee, Peter MJ Brown, John E Losey, Helen E Roy, and Rebecca Rice Smyth (2012) "Lessons from lady beetles: accuracy of monitoring data from US and UK citizen-science programs," *Frontiers in Ecology and the Environment*, 10 (9), 471–476.

Hamermesh, Daniel S. (2007) "Replication in economics," *Canadian Journal of Economics*, 40 (3), 715–733.

Hicks, Diana M (2007) "Global research competition affects measured US academic output," *Science and the University*, 223–242.

Himmelstein, Daniel and K Powell (2016) "The history of publishing delays," *Satoshi Village*, 10.

Hoppe-Wewetzer, Heidrun C, Georgios Katsenos, and Emre Ozdenoren (2019) "Experimentation, Learning, and Preemption," *Available at SSRN 3321231*.

Ioannidis, John P.A. (2005a) "Why most published research findings are false," *PLoS medicine*, 2 (8), e124.

——— (2005b) "Contradicted and initially stronger effects in highly cited clinical research," *Jama*, 294 (2), 218–228.

Ioannidis, John PA (2020) "Coronavirus disease 2019: the harms of exaggerated information and non-evidence-based measures," *European journal of clinical investigation*, 50 (4), e13222.

Ioannidis, John PA, Evangelia E Ntzani, Thomas A Trikalinos, and Despina G Contopoulos-Ioannidis (2001) "Replication validity of genetic association studies," *Nature genetics*, 29 (3), 306–309.

Iqbal, Shareen A, Joshua D Wallach, Muin J Khoury, Sheri D Schully, and John PA Ioannidis (2016) "Reproducible research practices and transparency across the biomedical literature," *PLoS biology*, 14 (1), e1002333.

Kisjes, I (2015) "The Evolution of Open Science–How Digitization is Transforming Research.[online] Elsevier."

Kosmala, Margaret, Andrea Wiggins, Alexandra Swanson, and Brooke Simmons (2016) "Assessing data quality in citizen science," *Frontiers in Ecology and the Environment*, 14 (10), 551–560.

Kupferschmidt, K (2020) “Study claiming new coronavirus can be transmitted by people without symptoms was flawed,” *Science*, 3.

Lazer, David MJ, Matthew A Baum, Yochai Benkler, Adam J Berinsky, Kelly M Greenhill, Filippo Menczer, Miriam J Metzger, Brendan Nyhan, Gordon Pennycook, David Rothschild et al. (2018) “The science of fake news,” *Science*, 359 (6380), 1094–1096.

Lukyanenko, Roman, Jeffrey Parsons, and Yolanda F. Wiersma (2016) “Emerging problems of data quality in citizen science,” *Conservation Biology*, 30 (3), 447–449.

Merton, Robert K (1957) “Priorities in scientific discovery: a chapter in the sociology of science,” *American sociological review*, 22 (6), 635–659.

——— (1969) “Behavior patterns of scientists,” *The American Scholar*, 197–225.

Michiels, Stefan, Serge Koscielny, and Catherine Hill (2005) “Prediction of cancer outcome with microarrays: a multiple random validation strategy,” *The Lancet*, 365 (9458), 488–492.

Nov, Oded, Ofer Arazy, and David Anderson (2014) “Scientists@ Home: what drives the quantity and quality of online citizen science participation?” *PloS one*, 9 (4), e90375.

Open Science Collaboration (2015) “Estimating the reproducibility of psychological science,” *Science*, 349 (6251), aac4716.

Palayew, Adam, Ole Norgaard, Kelly Safreed-Harmon, Tue Helms Andersen, Lauge Neimann Rasmussen, and Jeffrey V Lazarus (2020) “Pandemic publishing poses a new COVID-19 challenge,” *Nature Human Behaviour*, 4 (7), 666–669.

Pradhan, Prashant, Ashutosh Kumar Pandey, Akhilesh Mishra, Parul Gupta, Praveen Kumar Tripathi, Manoj Balakrishna Menon, James Gomes, Perumal

Vivekanandan, and Bishwajit Kundu (2020) “Uncanny similarity of unique inserts in the 2019-nCoV spike protein to HIV-1 gp120 and Gag,” *BioRxiv*.

da Silva, Jaime A Teixeira, Panagiotis Tsigaris, and Mohammadamin Erfanmanesh (2020) “Publishing volumes in major databases related to Covid-19,” *Scientometrics*, 1–12.

Soroye, Peter, Najeeba Ahmed, and Jeremy T Kerr (2018) “Opportunistic citizen science data transform understanding of species distributions, phenology, and diversity gradients for global change research,” *Global change biology*, 24 (11), 5281–5291.

Stephan, Paula E. (1996) “The economics of science,” *Journal of Economic literature*, 34 (3), 1199–1235.

Theobald, Ellinore J., Ailene K. Ettinger, Hillary K. Burgess, Lauren B. DeBey, Natalie R. Schmidt, Halley E. Froehlich, Christian Wagner, Janneke Hille Ris Lambers, Joshua Tewksbury, Melanie A. Harsch et al. (2015) “Global change and local solutions: Tapping the unrealized potential of citizen science for biodiversity research,” *Biological Conservation*, 181, 236–244.

Watson, David and Luciano Floridi (2018) “Crowdsourced science: sociotechnical epistemology in the e-research paradigm,” *Synthese*, 195 (2), 741–764.

Wynants, Laure, Ben Van Calster, Marc MJ Bonten, Gary S Collins, Thomas PA Debray, Maarten De Vos, Maria C Haller, Georg Heinze, Karel GM Moons, Richard D Riley et al. (2020) “Prediction models for diagnosis and prognosis of covid-19 infection: systematic review and critical appraisal,” *British Medical Journal*, 369.

Chapter 1

Refutation in Research: How to Improve Publication Quality

Abstract I examine how incentives for refutation affect publication quality. I build a sequential model of public experimentation with two scientists, a researcher and a refuter. The researcher chooses when to publish a result confirming a hypothesis, with a probability that the result has type I error. The publication quality corresponds to the probability that the hypothesis is valid, which is the complement of the probability of type I error. Once the result is published, the refuter starts experimenting but, unlike the researcher, he can choose between working on refutation or on an outside safe option at any time. When scientists are initially pessimistic about the hypothesis' confirmation and when refutation is costly for the researcher, the higher the refutation rewards, the lower the equilibrium publication quality in case the researcher is more efficient than the refuter. The opposite result holds when the refuter is more efficient. In an extension, when the researcher's experimentation is private, I prove that the publication quality is lower than in the public experimentation case, suggesting that transparency improves research quality.

Keywords: False positives, refutation, experimentation, publication quality.

JEL codes: C72, C91, O31.

1 Introduction

The provision of high-quality scientific research has received much attention in the last few decades, as noticed by The Economist (2013). Besides, according to Ioannidis (2005a), science is currently undergoing a "reproducibility crisis". This is particularly due to the high number of refutations. Refutation contrasts with replication, which consists in the repetition of a research study to determine if the findings can be applied to other data and methods. Hence, refutation occurs if the replication study's results are not statistically significant or contradict those of the initial study. As a matter of fact, many replication attempts failed. In cell biology, Begley and Ellis (2012) find that replication succeeds in only 11% of the attempted cases. In psychology, The Open Science Collaboration (2015), a conglomerate formed by scientists, shows that only half of the published results they consider are successfully replicated. These studies highlight that publication quality is a major issue and that both refutation and replication should be encouraged to increase research quality. Refutation issues concern also public health care: Ioannidis (2005b) shows that out of 45 medical studies from 1990-2004 claiming that an analyzed therapy is effective, only 44% are successfully replicated. More strikingly, in a study published in the *British Medical Journal*, Davis et al. (2017) examine cancer drugs approved by the European Medicines Agency (EMA) between 2009 and 2013. All new cancer drugs are given Europe-wide market approval through this agency. They find that 57% (39 drugs over 68) have no supporting evidence of better survival or quality of life when they enter the market. Moreover, a significant minority of drugs (12%) are approved solely on the basis of a single arm study. These findings question the quality of research conducted before a drug's approval, in addition to the paucity of studies conducted on each drug.

Hence, the debate about the reproducibility of research involves not only the scientific world, but also firms, policymakers and the public opinion. Sadly, Freedman et al. (2015) claims that every year, 28 billion dollars are spent in the US on preclinical research which is not reproducible. Besides, refutation is not a practice encouraged by academics and journals. Indeed, incentives are missing for researchers and journals: research prizes such as the Nobel prize reward creative and innovative research (but neither refutation nor replication), and journals competing for novelty have only little interest in refutations. As a matter of fact, Mueller-Langer et al. (2019) show that between 1974 and 2014, 0.1% of publications in the top 50 economics journals were replication studies. As emphasized by Hamermesh (2007), "the rewards to within-study scientific replication are small". In addition, refutation is not easy to implement in practice. Data and methods used by original papers are not always available and for some studies conducting experiments, replication attempts can be very costly. Refutation also implies challenging a peer. Mentioning a heated discussion caused by a replication attempt in epidemiology, Humphreys (2015) deplores that it can "contribute to the polarization of [a] discussion, make it hard for authors to acknowledge errors, and inhibit learning from this kind of replication exercise". Even though it is argued that journals should encourage replication by enforcing standards, scientists themselves, as Bruce Alberts, then editor of *Science*, emphasized the "need to develop a value system where simply moving on from one's mistakes without publicly acknowledging them severely damages, rather than protects, a scientific reputation".

Hopefully, many efforts have been done among the scientific community to encourage the research reproducibility. For instance, the Berkeley Initiative for Transparency in the Social Sciences (BITSS) was created in 2012 to reward transparent practices through the Leamer-Rosenthal Prizes. Furthermore, numerous journals, like the *Transactions on Replication Research* (TRR), have signed the Transparency and Openness Promotion (TOP) guidelines. The Center for Open Science created it in 2015 to increase transparency standards. Moreover, several journals created a section fully dedicated to replication, such as the *Journal of Experimental Psychology: General* and the *Journal of Applied Econometrics*. In addition, the International Initiative for Impact Evaluation have launched a replication programme to "improve the quality and reliability of impact evaluation evidence used for development decision-making". Besides, the *International Journal for Re-Views in Empirical Economics* (IREE) was initiated in 2017 to publish only replication studies. Step by step, publication quality is strenghened through all these advances.

To shed light on these issues, I study in this chapter a model in which the publication quality depends on refutation incentives, represented as rewards for refutation's success. In my model, there exists a hypothesis which can be either valid or invalid. Two scientists are uncertain about the hypothesis' validity and can conduct experiments to check it. At the beginning of the game, they have the same initial belief that the hypothesis is valid. They experiment sequentially. The first one, the researcher (she), starts working at time 0. While experimenting, she can observe for free some bad news which perfectly reveal that the hypothesis is invalid. Bad news are better suited than good news as there always remains some uncertainty when studying a hypothesis' validity. This bad news' arrival is public and exponentially distributed conditionally on the hypothesis being invalid. The fact that the bad news' arrival is public means that it is completely verifiable and every observation is reported on a data base which is available to the other scientist. This assumption is realistic as more and more journals ask for the provision of data bases and tools used for the published study.

At any time, she can publish a result confirming the hypothesis. There is a full publication bias towards the hypothesis validity. This is a simplifying assumption of the fact that results which prove a relationship between two phenomena are more likely to be published. As soon as the result is published, she obtains a flow payoff forever, which is consistent with Smaldino and McElreath (2016), who also study the refutation process. If, later, the hypothesis if refuted, she suffers a flow cost forever, which is higher than the reputation payoff obtained from publication. In that way, my model is consistent with Petersen et al. (2014), who empirically provide evidence that a researcher's reputation can decrease if she publishes invalidated science. Before publication, if no bad news has arrived yet, the researcher's belief that the hypothesis is valid increases deterministically. At the result's publication time, this belief is called the "publication quality". This quality is increasing in the speed at which bad news arrive. Accordingly, this speed is called the "researcher's speed". After publication, the researcher is no longer active and the second scientist, the refuter (he), starts working. At any time, he chooses between experimenting and working on an outside safe option. In this way, he differs from the researcher, who has no outside option and is enthusiastic to experiment first. The learning process is similar to the researcher and his learning speed is called the "refuter's speed". Furthermore, his bad news' arrival is public. If he receives some bad news, he can immediately publish his refutation of the researcher's work and enjoy a flow refutation payoff forever. The probability that the hypothesis is valid at some time after publication is called the "result reliability".

This model applies to two kinds of framework. The first one is academic competition. Working on an empirical study, the researcher would like to prove that some hypothesis is valid but is uncertain about it. The more she works on the data, the more she is sure that the hypothesis is valid, if this is the case. The publication bias towards the hypothesis validity implies that the researcher never finds in her interest to publish a result stating the hypothesis is invalid. Since she must report every details about the result, she cannot falsify the result. This assumption is realistic as several journals now require to provide data bases and tools used for a study. After the publication of the result, another scientist might work on refuting the hypothesis if he thinks his chances of success are high enough. The refutation work is published only if refutation is a success, as publishing a study confirming the published result is not attractive to journals. The second framework is drug testing. The researcher is a firm that is interested in proving a drug's efficiency. The hypothesis validity means that the drug is efficient; its invalidity means that is is not or that it induces dangerous side effects. The firm conducts an experiment on some patients whose health is evolving over time. These experiment's reports are public and the firm cannot cheat on it. If, at some point, their health suddenly deteriorates, that is, if the firm receives bad news, the drug cannot be commercialized. In this case, the firm earns no profit. Otherwise, the drug is commercialized with some probability that it is efficient. After the drug is launched on the market, a consumer association might study the drug's efficiency if she thinks the drug is probably dangerous.

In a first part, I study the strategic solution using backward induction. The "switching quality" is the equilibrium quality level at which the refuter, becoming too pessimistic about his refutation success, switches to the safe option. It increases with refutation rewards and decreases with the safe option payoff. Besides, it increases with the refuter's speed, reflecting that information is always valuable to the refuter. Taking into account the refuter's strategy, I then study the equilibrium publication quality. It depends on a threshold initial belief and a threshold refutation cost. Four cases exist: 1) The initial belief is so high that no matter how high the refutation cost is, the researcher immediately publishes the result and the refuter directly switches to the safe option; in this case, none of the scientists experiment; 2) The initial belief is low but the refutation cost is high: in this case, the researcher fears refutation and fully deters the refuter from experimenting: at the equilibrium, she publishes the result at the switching quality; there is no refutation risk and the refuter switches to the safe option right after the result's publication; hence, initial research findings are never challenged;[1] 3) The initial belief is high and the refutation cost low: the researcher does not fear refutation enough and immediately publishes the result; then the refuter experiments during a positive amount of time and there is a refutation risk; 4) The initial belief is low and the refutation cost is low; in this case, refutation fear is low but at the same time the researcher is not very optimistic when she starts experimenting: at the equilibrium, she experiments during a positive amount of time but publishes before reaching the switching quality; there is a refutation risk and the refuter experiments during a positive amount of time. In the latter case, I am able to study how the equilibrium publication quality varies with respect to the scientists' parameters. Interestingly, the effect of refutation rewards or the safe option payoff on the equilibrium publication quality depends on the difference of speed between the researcher and the refuter. When the researcher is more efficient, it decreases with refutation rewards and increases with the safe option payoff, meaning that an increase in the switching quality decreases the publication quality. The intuition for this result is that, as the switching quality increases, the refutation risk increases but stays low. Hence, the researcher prefers to enjoy her publication payoff earlier, at a lower publication quality. This result actually contradicts the literature which criticizes the paucity of refutation remuneration, as Hamermesh (2007) stated. When the refuter is more efficient, the equilibrium publication quality increases with refutation rewards and decreases with the safe option payoff. In this case, as the switching quality increases, the refutation risk is so high that the researcher prefers to publish the result later, at

[1]The researcher remains the only scientist to work on the hypothesis. She therefore uses entry deterrence to discourage the refutor from experimenting. Entry deterrence, which was mostly developped in the industrial organization literature, and more specifically by Salop (1979), occurs when a firm invests to deter entry from a potential competitor.

a higher publication quality. This result is in line with Hamermesh (2007). When the scientists are equally efficient, the equilibrium publication quality does not depend on the refutation rewards nor on the safe option payoff.

In a second part, I relax the full observability assumption, which means that the researcher's bad news arrival is private. In this way, I allow the researcher to fraud and publish a result she knows to be wrong. An "informed" type denotes a researcher who has received bad news and an "uninformed" type a researcher who has not received any bad news yet. I look for a perfect Bayesian equibrium in pure strategies, in which deviations from the equilibrium time are assumed to come from an informed type. An equilibrium in pure strategy does not always exist. Due to the participation constraint of an informed type, there exists no separating equilibrium in pure strategies. In this way, an informed researcher always has incentives to perfectly mimick an uninformed one by choosing the same publication time. Hence, I look for a pooling equilibrium. At equilibrium, if the initial belief is above the switching quality, the refuter never experiments and switches directly to the safe option. Otherwise, the refuter's experimentation span is fixed. In this way, the researcher can never prevent the refuter from experimenting. When the refutation cost is high, the incentive constraint of an informed type is not satisfied. Hence, there exists no pooling equilibrium in pure strategies. When the refutation cost is low, the participation and incentive constraints for both types are satisfied for any publication time. The pooling equilibrium which maximizes the uninformed researcher's payoff is to publish the result at time 0. This result suggests that transparency is better for research quality.

Related Literature. To begin with, this chapter is in line with the economics of science, a research field which Stephan (1996) explores in details, and more particularly with the research process literature. Bobtcheff et al. (2017) study researchers' incentives to publish their work before achieving maturity by fear of being preempted by the others. Applied to an hypothesis' testing, this maturity represents a small type I error. In this way, fighting for priority deteriorates the research quality. In my model, I do not study the impact of competition but rather the impact of refutation on publication quality. Besides, many recent papers focus on the worrying issue of type I error. The closest paper to this chapter is Kiri et al. (2018) in which one researcher must exert high level of effort in order to have a certain publication quality level. One colleague then decides whether to verify its quality. By studying Nash equilibria, they prove that incentivizing research effort and verification activities improves the expected quality of research. In my model, this result is similar or goes in the opposite direction depending on the difference of speed between the two scientists. The main differences are that, in my model, improving the expected publication quality is costly through the discounting effect and that this quality is known by the refuter, an assumption I relax in an extension. Besides, two other papers study how replication and refutation impact the research quality. Smaldino and McElreath (2016) theoretically show that replication and refutation slow but do not stop the poor research quality's flow. They assume that both replication and refutation are rewarded and that refutation is costly to the researcher, which is consistent with my model. By contrast, I focus on refutation. In this way, their results go in the same direction as mine. The second paper which studies replication is Maniadis et al. (2014), which provides a simple model and its testing. They find that a few independent replications dramatically increase the chances that the original finding is true. In contrast, in my model, I study refutation incentives rather than replication ones.

On the same topic, Libgober (2018) studies the impact of transparency on research quality. Transparency implies that no falsification is possible for researchers. He shows that imposing full transparency when the experiment is costly does not increase the publication quality. In my model, it increases the publication quality when experimenting is costless. Moreover, Yoder (2018), imposes full transparency and costly information acquisition for the researcher. He uses the terminology "positive results" for results which are statistically significant and "negative results" for those which are not. He provides a Bayesian framework in which negative results should be rewarded at the optimum so that research quality increases. In my model, I impose full transparency, as Yoder (2018), but information acquisition is costless and there is full publication bias towards positive result, which is a result in favor of hypothesis' validity.

Furthermore, in the economics of science, the publication bias issue is a topic close to this chapter. Indeed, as the literature emphasises, journals are more likely to publish papers with statistically significant results, which creates bias towards the confirmation of the tested hypothesis, as mentioned by Brodeur et al. (2016). Furukawa (2017) provides a model and some evidence to explain why publication bias is socially optimal whereas Andrews and Kasy (2019) provide econometric methods to identify publication bias. They find that adjusting for it substantially changes the number of significant results for economic laboratory experiments. On the other hand, this publication bias creates incentives for the researcher to perform adaptive selection, in the sense that she chooses which model to use, which hypothesis to test or which parameter to estimate after seeing the data. In this chapter, this publication bias is taken into account and results which proves the hypothesis invalidity cannot be published.

Besides, my model uses a strategic experimentation framework which Keller et al. (2005) develop in their seminal paper. In their model, though, many players experiment simultaneously whereas in my model, scientists experiment sequentially. Besides, Ozdenoren et al.(2019) study a model of experimentation and learning where the researchers share or keep private the information about the outcomes of their experiments. They show that information sharing can lead to more experimentation and higher welfare. In my model, I obtain similar results: the researcher experiments more and the publication quality is higher when transparency is imposed. Moreover, McClellan (2017) uses an experimentation model to study FDA decision rules. In his paper, a principal designs a mechanism to incentivize the agent's experimentation without monetary transfers. He shows that longer experimentation leads to more type I error, which is due to the fact that the principal keeps incentivizing the agent to experiment when the belief about the drug's efficiency is low. In his model, the learning process is Brownian and experimentation costly for the agent whereas in mine, it is a Poisson process and experimenting is costly through the discounting effect. Besides, I use refutation instead of a mechanism design to incentivize experimentation. Moscarini and Smith (2001) study an experimentation framework where a decision maker has to invest or not into costly R&D. They show that experimentation is increasing in beliefs about the product's quality. In my model, the latter assertion is true as the expected researcher's payoff increases in beliefs.

Lastly, my model is related to the persuasion literature. Di Tillio et al. (2017) study a model in which a researcher strategically manipulates an experiment to induce an evaluator to adopt a hypothesis. They find that in some cases, the experiment's manipulation benefits the evaluator. Moreover, Dur and Swank (2005) examine the selection of information collecting agents by policy makers. They show that unbiased advisers put highest effort in collecting information. In contrast, this chapter studies the case of a biased researcher who can manipulate information but instead of an evaluator taking a decision, a refuter may check the experiment's outcomes. Unlike Di Tillio et al. (2017), the experiment's manipulation does not benefit the refuter. In addition, Felgenhauer and Loerke (2017) analyze a Bayesian persuasion game in which the sender sequentially privately experiments to provide hard evidence. They find that the receiver obtains access to higher quality information than under public experimentation. My results go in the same direction in the sense that the publication quality and the researcher's experimentation are higher under public experimentation. Finally, Henry and Ottaviani (2019) study a model where an informer sequentially collects information through costly research to persuade an evaluator to implement a project. They show that giving authority to the informer is socially optimal when information acquisition is sufficiently costly. In my model, the researcher has some power in the sense that she is the only one to work on the hypothesis at the beginning of the game. However, her experimentation is checked by the refuter.

The chapter is organized as follows: Section 2 presents the model. In Section 3, the equilibrium of the game is analyzed and in Section 4, the effects of the parameters on the equilibrium publication quality are studied. In Section 5, an extension is provided where the researcher's signal is private. Section 6 concludes.

2 The Model

Time $t \in [0, \infty)$ is continuous and the discount rate is $r > 0$. There exists a hypothesis which can either be valid or invalid. The validity of the hypothesis is denoted H and its invalidity $\overline{H}$. Two scientists are uncertain about which event occurs and can conduct experiments to check it. At time $t = 0$, they have the same initial belief that the hypothesis is valid, $p_0 = P(H)$. They experiment sequentially.

2.1 The Researcher

The first scientist, *the researcher* (she), starts working at time 0. While experimenting, she can observe for free some bad news which perfectly reveals $\overline{H}$. This bad news' arrival is public and exponentially distributed with parameter $\lambda_R > 0$ conditionally on

$\overline{H}$, and 0 conditionally on H.[2] The fact that the bad news' arrival is public means that the experiment conducted by the researcher is fully transparent: every observation is reported on a data base, which, after publication, is available to the other scientist.

At any time, the researcher can publish a result confirming H. There is a full publication bias towards results confirming H: the researcher cannot publish a result stating $\overline{H}$. As soon as the result is published, she obtains a flow payoff $w_R > 0$ forever.[3] But if, at a later point in time, the hypothesis is refuted, she suffers a flow cost $c_R > 0$ from the refutation time onwards.[4]

Before publication, if no bad news has arrived yet, the researcher's belief that the hypothesis is valid increases deterministically. The total probability of observing no bad news before publication time t_R is $p_0 + (1 - p_0)e^{-\lambda_R t_R}$. The *publication quality* at publication time t_R, $\boldsymbol{p_R}(t_R)$, is the probability of H at the publication time, conditional on no bad news being observed up to that time; that is:

$$\boldsymbol{p_R}(t_R) = P(H| \text{ no bad news before } t_R) = \frac{p_0}{p_0 + (1 - p_0)e^{-\lambda_R t_R}}. \tag{1}$$

For a given publication time, this quality is increasing in λ_R. Accordingly, λ_R is called the *researcher's speed*. As λ_R increases, the researcher becomes more efficient at testing the hypothesis. The publication quality is strictly increasing in the amount of time spent experimenting. After publication, the researcher is no longer active and the second scientist can start working.

[2] The private case is treated in Section V.

[3] Experimentation is costly for the researcher only in the sense that time is discounted. Modelling an outside option for the researcher would complicate the analysis without changing the results. She would exert the outside option only when she is too pessimistic at time 0 or in case of a bad news' arrival. Consequently, when the initial belief is too low, she would switch directly to the safe option and never publish the result. In case of publication, she would also publish the result later, at a higher publication quality, as she has some security net in case of bad news.

[4] In this sense, my model is coherent with a publication payoff which would depend on the publication quality.

2.2 The Refuter

Once the result is published, the second scientist, *the refuter* (he), can start experimenting. At any time, he has one indivisible unit of resource to allocate between refutation and an outside safe option, which brings a flow payoff $s_F > 0$.[5]

Like the researcher, the refuter can observe for free some bad news which perfectly reveals $\overline{H}$. This bad news' arrival is public and exponentially distributed with parameter $\lambda_F > 0$ conditionally on $\overline{H}$, and 0 conditionally on H. In analogy with λ_R, λ_F is called *the refuter's speed*. The fact that the signal is public means that it is verifiable. In this way, a refuter cannot cheat by claiming he received some bad news.[6] If the refuter receives bad news, he can immediately publish his refutation of the researcher's work and enjoy a flow refutation payoff $w_F > 0$ forever.[7]

By Bayes' rule, if the refuter has kept experimenting until time t, his belief that H is valid, when no bad news has arised at time t, is:

$$\boldsymbol{p_F}(t, t_R) = P(H|\text{ no bad news before } t) = \frac{\boldsymbol{p_R}(t_R)}{\boldsymbol{p_R}(t_R) + \Big[1 - \boldsymbol{p_R}(t_R)\Big] e^{-\lambda_F(t-t_R)}}. \qquad (2)$$

$\boldsymbol{p_F}(t, t_R)$ is called *the result reliability* at time t when the result has been published at time t_R. For a given publication time, the result reliability is strictly increasing in t.

[5]There is an asymmetry between the researcher and the refuter. For the case where the researcher has an outside option, please refer to footnote 3.

[6]This would be the case if the bad news' arrival was private. Indeed, the refuter would have an incentive to lie to enjoy the refutation payoff as soon as possible. The result reliability would be uncertain.

[7]Allowing the refuter to replicate the result would have the same impact on the refuter's experimentation than his safe outside option. Indeed, the refuter would publish the replication result at some time only if she did not receive any bad news before that time. In this way, his strategy remains unchanged. Besides, it would complicate the researcher's strategy without changing the results by introducing an additional payoff in case of replication.

Assumption 1. The following holds:

$$c_R > \frac{r + \lambda_F}{\lambda_F} w_R, \tag{3}$$

$$w_F > \frac{r + \lambda_F}{\lambda_F} s_F. \tag{4}$$

Inequality (3) states that refutation is costly enough so that the researcher considers experimenting before publishing the result, rather than always publishing the result at time 0. Besides, it implies that if an unfrauded result is refuted, the researcher receives a negative flow payoff $w_R - c_R < -\frac{r}{\lambda_F} w_R$, that is, the overall impact of the publication on her payoff after refutation is negative. Inequality (4) states that working on refutation is attracting enough so that the refuter does not always choose to switch directly to his outside safe option. It implies that a refuter which is certain of $\overline{H}$ works forever on refutation.

Figure 1 shows the evolution of public beliefs $\{p_t\}_{t \geq 0}$ if no scientist observes any signal, the refuter switches only once to the outside option and the refuter's speed is higher than the researcher's speed. This evolution is Markovian with a change of regime at t_R. The probability that the hypothesis is valid increases up to the belief at which the researcher publishes (the publication quality), then increases again up to the belief at which the refuter switches to the outside option; after both scientists have stopped experimenting, the result reliability remains constant.

2.3 Explanation of the difference between the scientists' speeds

The researcher can be more or less efficient than the refuter. The former being less efficient than the latter can reflect the term "dwarfs on the shoulder of giants". It takes longer for pioneers to work on a project than followers, who can benefit about the work which has been done on it. For example, in 1994, and after more than three centuries

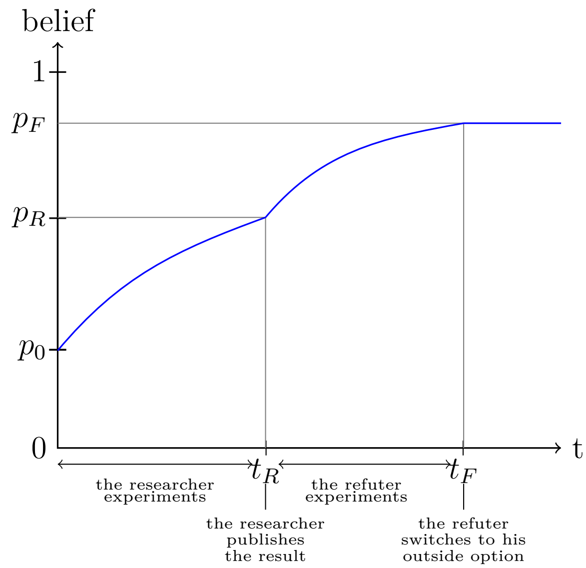

Figure 1 – The evolution of belief if no scientist observes any signal and the refuter switches only once to the outside option.

after its statement, the mathematician Andrew Wiles succeeded in proving the Fermat theorem, even many mathematicians made progress on it in between.

By contrast, the researcher being more efficient than the refuter could reflect the well widespread division of work. The most efficient scientists work on new ideas whereas the least efficient ones work on old ideas, such as refutation.

3 Equilibrium Strategies

By backward induction, I start by studying the refuter's strategy.

3.1 The Refuter's Equilibrium Strategy

The refuter has no incentive to switch more than once from experimenting to his safe option as his belief remains unchanged while working on the safe option and his payoffs are discounted. His expected payoff at publication time t_R, if he switches to his safe option at time $t_F \geq t_R$, is:

$$W_F(t_R, t_F) = \boldsymbol{p_R}(t_R) \int_{t_F}^{+\infty} s_F e^{-r(t-t_R)} dt + [1 - \boldsymbol{p_R}(t_R)] E_{t_R} \left[\mathbb{1}_{\tilde{t}<t_F} \int_{\tilde{t}}^{+\infty} w_F e^{-r(t-t_R)} dt + \mathbb{1}_{\tilde{t}>t_F} \int_{t_F}^{+\infty} s_F e^{-r(t-t_R)} dt \right]. \quad (5)$$

where $\tilde{t}$ is the random time at which the refuter first observes bad news. This payoff is the sum of two terms. First, with probability $\boldsymbol{p_R}(t_R)$, the hypothesis is valid and the refuter never receives bad news. Thus, he keeps experimenting until reaching time t_F, at which time he switches to the safe option and thereafter enjoys the safe payoff s_F. Second, with probability $1 - \boldsymbol{p_R}(t_R)$, the hypothesis is invalid. If the refuter observes a signal before switching to the safe option ($\tilde{t} \leq t_F$), he enjoys the refutation payoff from there on. Otherwise ($\tilde{t} > t_F$), he switches to the safe option at time t_F and enjoys thereafter the safe option payoff. Denote p_R the publication quality and p_F the result reliability which are given by

$$p_R = \boldsymbol{p_R}(t_R),$$
$$p_F = \boldsymbol{p_F}(t_F, t_R).$$

For each publication time t_R and switching time t_F, there exist a unique publication quality p_R and a unique result reliability p_F. In this way, the refuter's payoff can be expressed as a function of p_R and p_F only. Hence, the refuter's payoff can be rewritten as:

$$\tilde{W}_F(p_R, p_F) = \frac{1}{r} \left[(1 - p_R) \frac{\lambda_F}{r + \lambda_F} w_F + \frac{p_R}{p_F} \left(\frac{p_R}{1 - p_R} \frac{1 - p_F}{p_F} \right)^{\frac{r}{\lambda_F}} \left(s_F - \frac{\lambda_F}{r + \lambda_F} (1 - p_F) w_F \right) \right]$$

I call the equilibrium quality at which the refuter switches to the safe option *the switching quality.* The refuter's equilibrium strategy is stated in the following proposition.

Proposition 1. *The refuter's equilibrium switching quality p_F^* is:*

$$p_F^* \equiv 1 - \frac{r}{\lambda_F} \frac{s_F}{w_F - s_F} \in (0, 1).$$

- *If $p_F < p_F^*$, the refuter keeps experimenting;*
- *If $p_F \geq p_F^*$, he switches to the safe option.*

Interestingly, the switching quality does not depend on the publication quality. This is due to the Markovian evolution of belief. In this way, the refuter always ensures that the result reliability is at least p_F^*, regardless of the researcher's strategy. If the publication quality is already equal to or higher than the switching quality, the refuter never experiments and switches to the safe option right after the result's publication. Otherwise, the refuter experiments a positive amount of time. Besides, the switching quality increases with the refutation payoff. Indeed, as the refutation payoff increases, the expected discounted payoff of experimenting increases. Therefore, the refuter can afford being more optimistic and experiments longer, leading to a higher switching quality. A similar reasoning explains why the switching quality decreases with the safe option payoff. Moreover, this switching quality increases in the refuter's speed. As the refuter becomes more efficient, he can experiment more and trigger publication for a higher belief's threshold. Lastly, the switching quality decreases in the discount factor. Indeed, as the discount effect becomes stronger, the refuter's expected discounted payoff decreases.

Besides, the refuter's experimentation time Δt^*, as a function of the researcher's publication quality, is

$$\Delta t^* = t_F^* - t_R = \frac{1}{\lambda_F} \Big[\log \Big(\frac{1 - p_R}{p_R} \Big) + \log \Big(\frac{p_F^*}{1 - p_F^*} \Big) \Big]. \tag{6}$$

For a given switching quality, it decreases in the publication quality. As a matter of fact, as the publication quality increases, the refuter becomes more pessimistic about his chances of success and experiments less. The effect of the scientists' parameters on the refuter's experimentation time depends on their effect on the publication quality, which will be studied in Section IV.

3.2 The Researcher's Equilibrium Strategy

If the initial belief is above the switching quality, the researcher knows the refuter immediately switches to the safe option. Hence, in equilibrium, she publishes the result at time 0. Otherwise, that is, if $p_0 < p_F^*$, there exists a refutation risk and the researcher faces the following trade-off. i) By publishing exactly at the switching quality, she fully prevents the refuter from experimenting as the latter switches directly to the safe option. In this case, she enjoys her publication payoff with no refutation risk. ii) By publishing before reaching the switching quality, she faces a refutation risk but enjoys her publication payoff earlier. In this case, her expected payoff at time 0 is:

$$W_R(t_R, t_F) = e^{-rt_R}\left(p_0+(1-p_0)e^{-\lambda_R t_R}\right)E_{t_R}\left[\int_{t_R}^{+\infty} w_R e^{-r(t-t_R)}dt-(1-\boldsymbol{p_R}(t_R))\mathbb{1}_{\tilde{t}<t_F}\int_{\tilde{t}}^{+\infty} c_R e^{-r(t-t_R)}dt\right]. \tag{7}$$

The part under brackets is the expected difference between two terms: the discounted publication payoff and the discounted refutation cost in case the hypothesis is invalid and refutation happens before the refuter switches to the safe option. The payoff is discounted at time 0 and occurs only if the researcher does not observe any signal before publication. In case she receives some bad news, her expected payoff is 0 as she cannot publish the result. The researcher's srategy can be expressed in terms of both the publication quality and the switching quality only. The equilibrium publication quality

p_R^* satisfies:

$$p_R^* = \text{Arg} \max_{p_R \leq p_F^*} \tilde{W}_R(p_R, p_F^*),$$

where for all $p_R \in [p_0, p_F^*]$,

$$\tilde{W}_R(p_R, p_F^*) = \frac{p_0}{p_R}\left(\frac{p_0}{1-p_0}\frac{1-p_R}{p_R}\right)^{\frac{r}{\lambda_R}}\left[\frac{w_R}{r} - (1-p_R)\frac{\lambda_F}{r+\lambda_F}\frac{c_R}{r}\left(1-\left(\frac{1-p_F^*}{p_F^*}\frac{p_R}{1-p_R}\right)^{\frac{r+\lambda_F}{\lambda_F}}\right)\right]. \tag{8}$$

Detailed computation can be found in the appendix. This function is continuous at p_F^*. There is no closed form solution for the equilibrium publication quality, but four cases can be distinguished, depending on two thresholds, $\bar{c}_R$ and $\bar{p}_0$, defined as follows:

$$\bar{c}_R \equiv \left(1 + \frac{\lambda_F}{\lambda_R}\frac{w_F - s_F}{s_F}\right)w_R, \tag{9}$$

$$\bar{p}_0 \text{ such that } \left.\frac{\partial \tilde{W}_R(p_R, p_F^*)}{\partial p_R}\right|_{p_R = \bar{p}_0} = 0. \tag{10}$$

$\bar{c}_R$ represents the threshold cost above which refutation is too costly for the researcher. In this case, she fully prevents the refuter from experimenting by publishing the result at the switching quality. The threshold $\bar{p}_0$ represents the quality at which the researcher's payoff starts decreasing. If the initial belief is above this threshold, experimenting before publishing the result decreases the researcher's payoff. In this way, the latter does not wait and publishes the result at time 0.

Proposition 2. *The equilibrium publication quality p_R^* splits in four cases:*

- Optimistic certain publication: *if $p_0 \geq p_F^*$, the researcher publishes at time 0, that is, $p_R^* = p_0$;*

- Pessimistic certain publication: *if $c_R \geq \bar{c}_R$ and $p_0 < p_F^*$, there is full deterrence of the refuter's experimentation, that is, $p_R^* = p_F^*$;*

- Pessimistic risky publication: *if* $c_R < \overline{c}_R$ *and* $p_0 < \overline{p}_0$*, the researcher publishes before reaching the switching quality, that is,* $p^*_R < p^*_F$*;*

- Optimistic risky publication: *If* $c_R < \overline{c}_R$ *and* $p^*_F > p_0 \geq \overline{p}_0$*, the researcher publishes at time* 0*, that is,* $p^*_R = p_0$.

If the initial belief is above the switching quality, the refuter never experiments and switches directly to the safe option. There is no refutation risk and the researcher publishes the result at time 0. This is the "optimistic certain publication". When the refutation cost is above the threshold, refutation risk is too costly for the researcher so that she fully discourages the refuter from experimenting and publishes at the switching quality: this case is called "pessimistic certain publication". Otherwise, the refutation risk is small and the researcher optimally chooses not to prevent the refuter from experimenting. Her choice depends on the initial belief. If p_0 is above $\overline{p}_0$, experimenting only decreases her payoff, hence, she publishes at time 0 even if there is a refutation risk: this is the "optimistic risky publication" case. Otherwise, that is, if the refutation cost is below the threshold and the initial belief low, at the equilibrium, the researcher publishes before reaching the switching quality. There exists a refutation risk. This last case is called "pessimistic risky publication".

These four cases are represented in the graph (c_R, p_0), as shown in Figure 2. Figure 2 represents the case where λ_R is smaller than a threshold $\overline{\lambda}_R$ defined in the appendix. In this case, when the cost is smaller than $\overline{c}_R$, the researcher fears refutation: her tradeoff is between publishing before reaching the switching quality, which involves a refutation risk, or publishing at the switching quality, thus avoiding any refutation risk. If her speed is higher than $\overline{\lambda}_R$, $\overline{p}_0(\overline{c}_R)$ is lower than p^*_F. In this case, when the initial prior is just below the switching quality, the researcher does not fear refutation enough and her tradeoff is between publishing at time 0, with a refutation risk, and publishing at the switching quality. Detailed computations can be found in the appendix.

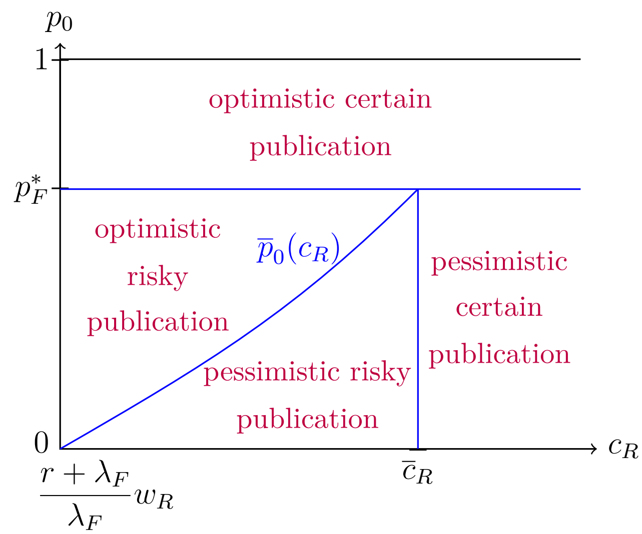

Figure 2 – The four publication cases on the graph (c_R, p_0)

The "optimistic certain publication" case occurs when, for instance, the researcher obtains some of the experiment's observations very fast and these are good enough in favor of $\overline{H}$ to publish immediately the result. This is what is called "interim data analysis" in research, as Landis et al. (2012) pointed out, which occurs when the analysis of the experiment's data is conducted before data collection has been completed. Hence, a researcher who is very optimistic about the experiment's result stops the experiment very quickly and publishes a result with incomplete experiment's observations. Interestingly, in case of "pessimistic risky publication", I am able to study the effect of the researcher and refuter's parameters on the equilibrium publication quality. The following section focuses on this case.

4 Pessimistic Risky Publication

In this case, $p_0 < \overline{p}_0$ and $c_R < \overline{c}_R$. The refutation cost is relatively low, but it must satisfy Inequality 3 in Assumption 1. I derive a necessary condition for this case to hold, which depends on a threshold $\overline{\lambda}_F$ such that:

$$\overline{\lambda}_F \equiv \sqrt{\frac{r\lambda_R s_F}{w_F - s_F}}.$$

Lemma 1. *A necessary condition for the existence of the pessimistic risky publication case is:*

$$\lambda_F > \overline{\lambda}_F. \tag{11}$$

This inequality provides the existence of the optimistic risky publication case as well. If the refuter's speed is below this threshold, the refuter is slow in learning, leading to a low switching quality. If the initial belief is below the switching quality, the refutation risk exists and it is already too costly for the researcher to publish before reaching the switching quality. If the initial belief above the switching quality, there is no refutation risk and the researcher publishes the result at time 0. The next proposition summarizes the comparative static analysis of the equilibrium publication quality, depending on the scientists' respective speeds.

Proposition 3. *The equilibrium publication quality p_R^*:*

- *decreases with the researcher's publication payoff w_R;*
- *increases with the researcher's refutation cost c_R;*
- *increases with the researcher's speed λ_R;*
- *decreases with the refuter's refutation payoff w_F if $\lambda_R > \lambda_F$ and increases with it otherwise;*
- *increases with the refuter's safe option payoff s_F if $\lambda_R > \lambda_F$ and decreases with it otherwise;*
- *increases with the refuter's speed λ_F.*

Firstly, the equilibrium publication quality is decreasing in the researcher's publication payoff. This effect is not surprising and similar to the one I obtained in the analyses of the refuter's strategy. As her publication payoff increases, the expected discounted publication payoff increases and the expected discounted cost remains the same. As a consequence, the researcher fears less refutation and publishes at a lower publication quality. Or, in terms of publication time, she publishes the result earlier. In this way, the more the researcher is remunerated, the less cautious she is about the publication quality. As a consequence, the researcher should not receive too much merit from publishing her result. Similarly, the equilibrium publication quality increases with the refutation cost.

Secondly, the equilibrium publication quality increases with the researcher's speed. As the researcher learns faster, she gains in experimenting longer to increase the amount of evidence in favour of the hypothesis validity, thus inducing a lower refutation risk. Hence, the equilibrium publication quality increases. Accordingly, the researcher's ability benefits the publication quality. The analysis in terms of publication time is similar to Bobtcheff and Levy (2017). The researcher's speed impacts the latter in two ways. As the researcher becomes more efficient, the value of experimenting longer increases. Meanwhile, she reaches a given publication quality faster. The first effect always dominates the second one. Accordingly, the publication time increases.

Thirdly, and more surprisingly, is the interplay between the refutation payoff and the equilibrium publication quality. This effect depends on the difference of speed between the researcher and the refuter. As the refuter's refutation payoff increases or his safe option payoff decreases, at stated in Proposition 1, the switching quality increases. When the researcher is more efficient than the refuter, the refutation threat is weak and the refuter's experimentation phase is poorly efficient. The researcher takes advantage on it by publishing earlier to enjoy her publication payoff sooner. Hence, the equilibrium publication quality decreases. When the refuter is more efficient, the refutation threat is strong and the researcher prefers to publish later to decrease the refutation risk. Consequently, the equilibrium publication quality increases. When the scientists' speeds are equal, the increase in the expected benefit is she publishes earlier exactly balances the decrease in the expected cost if she publishes later. Accordingly, the equilibrium publication quality remains unchanged. As a consequence, the policy implications are different whether the researcher is more or less efficient than the refuter. If the researcher is more efficient, the refuter should receive low refutation rewards or have a good outside option to benefit the publication quality. The policy implication actually contradicts the literature which criticizes the paucity of rewards for refutation, as mentioned by Hamermesh (2007). If the refuter is more efficient, he should receive high refutation rewards or have a poor outside option, which goes in the same direction as Hamermesh (2007). There is no interest in having both speeds equal as, in this case, refutation rewards has no impact on the publication quality.

Lastly, the equilibrium publication quality is increasing in the refuter's speed. An increase in the refuter's speed has two effects on the researcher's strategy. The indirect effect is an increase in the switching quality, which could a priori increase or decrease the publication quality, depending on the researcher's speed relative to the refuter's speed. The direct effect is that the refuter receives bad news at a higher speed, inducing a higher refutation risk. In this way, the equilibrium publication quality increases. It turns out that the direct effect dominates the indirect one. Hence, the equilibrium publication quality increases. In terms of time, some result reliability level is reached faster, so the refutation risk increases, which tends the researcher to experiment longer. The switching quality could a priori increase or decrease the publication time. The first effect always dominates the second one, which means that the researcher always experiments more. In this way, the refuter's speed always benefits the publication quality.

In light of these results, the effect of the researcher's parameters on the refuter's experimentation time can be derived through Equation 6 evaluated at t_R^*. The refuter's experimentation time increases in the publication payoff and decreases in both the refutation cost and the researcher's speed. Indeed, as the publication payoff decreases, the publication quality increases, inducing lower chances of refutation success. Consequently, the refuter experiments less. The analysis is similar if the refutation cost or the researcher's speed increase. Furthermore, an increase in the refutation payoff or a decrease in the refuter's safe option payoff on the refuter's equilibrium experimentation time splits two effects. The first and direct effect is positive through the increase of the equilibrium switching quality. The second and strategic effect depends on the evolution of the equilibrium publication quality. If the researcher is more efficient than the refuter, the equilibrium publication quality decreases. Accordingly, the strategic effect is positive and the refuter's equilibrium experimentation time increases. When the refuter is more efficient, the equilibrium publication quality increases. In this way, the strategic effect is negative and the total impact on the refuter's equilibrium experimentation time is uncertain. At last, the impact of the refuter's speed on the refuter's experimentation time is less clear than the latter impacts. In addition to the direct effect (through the switching quality) and the strategic effect (through the publication quality), there exists a third direct effect. As the refuter becomes more efficient, the refuter experiments less to reach some threshold. Consequently, the third effect is negative. The overall impact on the refuter's equilibrium experimentation time is unclear.

5 Unknown Publication Quality

In this section, the researcher's learning process is assumed to be private: she does not have to publicly report every observation of the experiment. In this way, she has an incentive to fraud by hiding the signal's arrival to publish her result and enjoy the publication payoff. Consequently, the publication quality is uncertain for the refuter. Lying is not costly for the researcher. A researcher who has not received any bad news before publication is called an *uninformed type* and one who has received an *informed type*. I look for a perfect Bayesian equilibrium in pure strategies, in which each type publishes at a unique time. I assume that if the researcher deviates, the refuter believes she is an informed type.[8] In this case, he works forever on refutation.

If a separating equilibrium exists, the participation constraint of an informed type should be satisfied. Due to Inequality 3, it is too costly for an informed type, who is perfectly identified by the refuter, to publish. Hence, her participation constraint is never satisfied. This leads to the following lemma.

Lemma 2. *There exists no separating equilibrium in pure strategies.*

The informed type has always interest in deviating to mimick an uninformed type. I therefore look for a pooling equilibrium in pure strategies.

5.1 The refuter's pooling equilibrium strategy

As the publication quality is uncertain, the refuter has no additional information compared to time 0. Let t_R and t_F be, respectively, the publication time and the refuter's switching time. By backward induction, the refuter maximizes his expected payoff at time 0:

$$W_F^P(t_R, t_F) = E_0\left[p_0 \int_{t_F}^{+\infty} s_F e^{-rt} dt + (1-p_0)\left(\mathbb{1}_{\tilde{t}<t_F} \int_{\tilde{t}}^{+\infty} w_F e^{-rt} dt + \mathbb{1}_{\tilde{t}>t_F} \int_{t_F}^{+\infty} s_F e^{-rt} dt\right)\right].$$

[8]This assumption is more restrictive than if the refuter believes a deviation is made by an uninformed type. If it was the case, both types would have more incentives to deviate and an equilibrium would exist under narrowed conditions.

His payoff is similar to Equation (5), but with a different starting belief. In the baseline model, there is a learning effect through the researcher's experimentation. Hence, the publication quality, $\boldsymbol{p_R}(t_R)$, is used as a starting belief. By contrast, here, there is no learning and the starting belief is p_0. Furthermore, in the baseline model, the equilibrium switching quality does not depend on the publication quality. As a consequence, in this extension, the switching quality remains the same and does not depend on p_0.

Proposition 4. *The refuter's strategy is the following:*

- *If $p_0 \geq p_F^*$, the refuter does not experiment and switches to the safe option right after the result's publication;*
- *If $p_0 < p_F^*$, the refuter experiments during a fixed amount of time Δt^{**} with*

$$\Delta t^{**} = t_F^{**} - t_R = \frac{1}{\lambda_F}\Big[\log\Big(\frac{1-p_0}{p_0}\Big) + \log\Big(\frac{p_F^*}{1-p_F^*}\Big)\Big].$$

If the initial belief is higher than the switching quality, there is no refutation risk and the researcher publishes the result at time 0. If the initial belief is lower than the switching quality, the refuter experiments during the time needed to reach the switching quality, starting from the initial belief. It is interesting to note the difference between the equilibrium refuter's experimentation times in the baseline model (Equation 6) and in this extension is the starting belief, which is p_0 instead of p_R. In this extension, the refutation risk does not depend on the publication quality. It implies that the researcher cannot prevent the refuter from experimenting, in contrast to the baseline model.

5.2 The researcher's pooling equilibrium strategy

If the initial belief is above the switching quality, there is no refutation risk. Consequently, the researcher publishes at time 0. This is the "optimistic certain publication" case.

Otherwise, a pooling equilibrium with publication time t_R must satisfy the participation and incentive constraints for an informed type, who has received bad news before time t_R, and for an uninformed type, which are

$$\begin{cases} W_R^i(t_R) & \geq 0, \\ W_R^i(t_R) & \geq W_R^i(t),\ \forall t \neq t_R, \\ W_R^u(t_R) & \geq 0, \\ W_R^u(t_R) & \geq W_R^u(t),\ \forall t \neq t_R, \end{cases}$$

where

$$W_R^i(t_R) = e^{-rt_R}(1-p_0)(1-e^{-\lambda_R t_R})E_{t_R}\left[\int_{t_R}^{+\infty} w_R e^{-r(x-t_R)}dx - \mathbb{1}_{\tilde{t}<t_R+\Delta t^{**}}\int_{\tilde{t}}^{+\infty} c_R e^{-r(x-t_R)}dx\right]$$

$$W_R^i(t) = e^{-rt}(1-p_0)(1-e^{-\lambda_R t})E_t\left[\int_t^{+\infty} w_R e^{-r(x-t)}dx - \int_{\tilde{t}}^{+\infty} c_R e^{-r(x-t)}dx\right]$$

$$W_R^u(t_R) = e^{-rt_R}\big(p_0+(1-p_0)e^{-\lambda_R t_R}\big)E_{t_R}\left[\int_{t_R}^{+\infty} w_R e^{-r(x-t_R)}dx - (1-\boldsymbol{p_R}(t_R))\mathbb{1}_{\tilde{t}<t_R+\Delta t^{**}}\int_{\tilde{t}}^{+\infty} c_R e^{-r(x-t_R)}dx\right]$$

$$W_R^u(t) = e^{-rt}\big(p_0+(1-p_0)e^{-\lambda_R t}\big)E_t\left[\int_t^{+\infty} w_R e^{-r(x-t)}dx - (1-\boldsymbol{p_R}(t))\int_{\tilde{t}}^{+\infty} c_R e^{-r(x-t)}dx\right]$$

If an informed type deviates, the refuter knows that the latter has received bad news and works forever on refutation. Due to Inequality (3), her payoff in this case, $W_R^i(t)$, is always negative, so that if the participation constraint of an informed type is satisfied, so is her incentive constraint. The former is satisfied as long as c_R is lower than some threshold $\hat{c}_R$, defined below.

$$\hat{c}_R = \frac{r+\lambda_F}{\lambda_F}\left[1-\left(\frac{p_0}{1-p_0}\frac{rs_F}{\lambda_F w_F-(r+\lambda_F)s_F}\right)^{\frac{r+\lambda_F}{\lambda_F}}\right]^{-1} w_R. \qquad (12)$$

If the refutation cost is above this threshold, it is too costly for an informed type to mimick an uninformed type. Hence, in this case, there exists no pooling equilibrium. If the refutation cost is below the threshold, the informed type's constraints are satisfied for any $t_R \in \mathbb{R}^+$.

The participation and incentive constraints for an uninformed type are always satisfied for $c_R \geq \hat{c}_R$. Consequently, a pooling equilibrium exists for any $t_R \in \mathbb{R}^+$. As the payoff for an uninformed researcher decreases in t_R, the pooling equilibrium which maximizes the uninformed researcher's payoff is $t_R = 0$. This corresponds to the "optimistic risky publication" case.

Proposition 5. *When $c_R > \hat{c}_R$, there exists no equilibrium in pure strategies. Otherwise, there exists an infinity of pooling equilibria. The pooling equilibrium which maximizes the uninformed researcher's payoff is the one where the researcher always publishes the result at time* 0.

The pooling equilibrium which maximizes the uninformed researcher's payoff is represented on the graph (p_0, c_R) in Figure 3. $\hat{p}_0(c_R)$ represents the initial belief for which the threshold cost $\hat{c}_R$ is achieved, that is, $\hat{p}_0$ solves the equation $\hat{c}_R^{-1}(p_0) = \hat{p}_0$. In this pooling equilibrium, the publication quality is always equal to the initial belief. Interestingly, the publication quality is greater in the baseline model, suggesting that transparency benefits the publication quality.

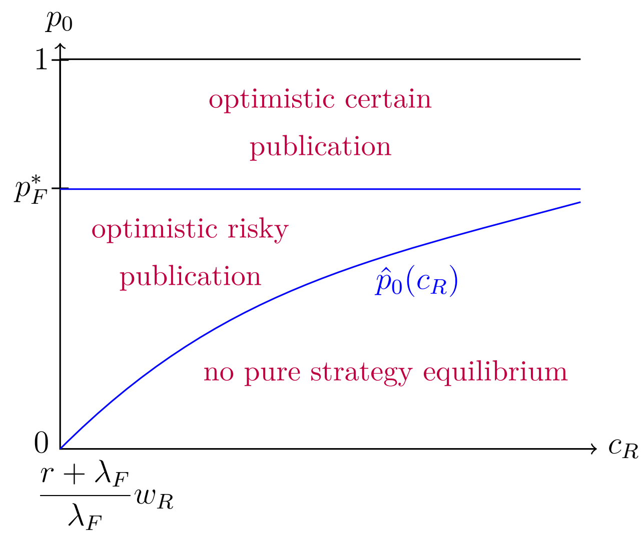

Figure 3 – The pooling equilibrium on the graph (c_R, p_0) where the uninformed researcher maximizes her payoff

6 Conclusion

Many studies have been devoted to replication and refutation but the impact of refutation's rewards on the publication quality has not been specifically analyzed. In this paper, I prove that the effect of refutation rewards on the publication quality depends on the difference of speed between the researcher and the refuter. If the researcher is more efficient, the effect is negative, which contradicts the literature's main policy recommendation, as Hamermesh (2007) stated. If the refuter is more efficient, the effect is positive, which is in line with Hamermesh (2007). I also give other policy recommendation regarding the researcher and the refuter's incentives. Interestingly, transparency always increases the publication quality in my model. To go further, it would be interesting to empirically test whether my theoretical predictions are confirmed in reality. For instance, one could use the number of citations as a measure of the publication quality, as Petersen et al. (2014) do. The researcher's publication payoff and refutation cost would be the impacts of the publication and the refutation on her total number of citations. An interesting topic for future research would be to analyze the journals' incentives to publish refutation results.

7 Appendix

Proof of Proposition 1

The refuter's expected payoff at publication time t_R is:

$$W_F(t_R, t_F) =$$
$$E_{t_R}\left[\boldsymbol{p_R}(t_R)\int_{t_F}^{+\infty} s_F e^{-r(t-t_R)}dt + (1-\boldsymbol{p_R}(t_R))\left(\mathbb{1}_{\tilde{t}<t_F}\int_{\tilde{t}}^{+\infty} w_F e^{-r(t-t_R)}dt + \mathbb{1}_{\tilde{t}>t_F}\int_{t_F}^{+\infty} s_F e^{-r(t-t_R)}dt\right)\right],$$

where $\tilde{t}$ follows a truncated exponential law on $[t_R, +\infty)$ of parameter λ_F;

$$= \boldsymbol{p_R}(t_R)\frac{1}{r}s_F e^{-r(t_F-t_R)} + (1-\boldsymbol{p_R}(t_R))\frac{1}{r}\left[w_F e^{(r+\lambda_F)t_R}\int_{t_R}^{t_F}\lambda_F e^{-(r+\lambda_F)t}dt + s_F e^{-r(t_F-t_R)}e^{\lambda_F t_R}\int_{t_F}^{+\infty}\lambda_F e^{-\lambda_F t}dt\right]$$
$$= \boldsymbol{p_R}(t_R)\frac{1}{r}s_F e^{-r(t_F-t_R)} + (1-\boldsymbol{p_R}(t_R))\frac{1}{r}\left[w_F\frac{\lambda_F}{r+\lambda_F}(1-e^{-(r+\lambda_F)(t_F-t_R)}) + s_F e^{-(r+\lambda_F)(t_F-t_R)}\right]$$
$$= \frac{1}{r}\left[(1-\boldsymbol{p_R}(t_R))\frac{\lambda_F}{r+\lambda_F}w_F + e^{-r(t_F-t_R)}\left(\boldsymbol{p_R}(t_R)s_F + e^{-\lambda_F(t_F-t_R)}(1-\boldsymbol{p_R}(t_R))\left(s_F - w_F\frac{\lambda_F}{r+\lambda_F}\right)\right)\right].$$

Moreover, by Equation 2, $W_F(.,.)$ can be expressed as a function of p_F and p_R with

$$\begin{cases} p_R = & \boldsymbol{p_R}(t_R) \\ p_F = & \boldsymbol{p_F}(t_F, t_R) \end{cases}$$
$$\Leftrightarrow \begin{cases} t_R = & \dfrac{1}{\lambda_R}\left(\log\left(\dfrac{1-p_0}{p_0}\right) + \log\left(\dfrac{p_R}{1-p_R}\right)\right) \\ t_F - t_R = & \dfrac{1}{\lambda_F}\left(\log\left(\dfrac{1-p_R}{p_R}\right) + \log\left(\dfrac{p_F}{1-p_F}\right)\right) \end{cases}$$
$$\Leftrightarrow \begin{cases} t_R = & \dfrac{1}{\lambda_R}\left(\log\left(\dfrac{1-p_0}{p_0}\right) + \log\left(\dfrac{p_R}{1-p_R}\right)\right) \\ t_F = & \dfrac{1}{\lambda_R}\log\left(\dfrac{1-p_0}{p_0}\right) + \dfrac{\lambda_R-\lambda_F}{\lambda_F\lambda_R}\log\left(\dfrac{1-p_R}{p_R}\right) + \dfrac{1}{\lambda_F}\log\left(\dfrac{p_F}{1-p_F}\right). \end{cases}$$

Hence, I have

$$\begin{aligned}\tilde{W}_F(p_R,p_F) &= W_F\left(\frac{1}{\lambda_R}\left(\log\big(\frac{1-p_0}{p_0}\big)+\log\big(\frac{p_R}{1-p_R}\big)\right),\frac{1}{\lambda_R}\log\big(\frac{1-p_0}{p_0}\big)+\frac{\lambda_R-\lambda_F}{\lambda_F\lambda_R}\log\big(\frac{1-p_R}{p_R}\big)+\frac{1}{\lambda_F}\log\big(\frac{p_F}{1-p_F}\big)\right)\\ &= \frac{1}{r}\left[(1-p_R)\frac{\lambda_F}{r+\lambda_F}w_F+\left(\frac{p_R}{1-p_R}\frac{1-p_F}{p_F}\right)^{\frac{r}{\lambda_F}}\left[p_Rs_F+p_R\frac{1-p_F}{p_F}\left(s_F-w_F\frac{\lambda_F}{r+\lambda_F}\right)\right]\right]\\ &= \frac{1}{r}\left[(1-p_R)\frac{\lambda_F}{r+\lambda_F}w_F+\frac{p_R}{p_F}\left(\frac{p_R}{1-p_R}\frac{1-p_F}{p_F}\right)^{\frac{r}{\lambda_F}}\left(s_F-\frac{\lambda_F}{r+\lambda_F}(1-p_F)w_F\right)\right].\end{aligned}$$

The switching quality p_F^* is such that

$$\begin{aligned}p_F^* &= \arg\max_{p_F\in[p_R,1]}\tilde{W}_F(p_R,p_F)\\ &= 1-\frac{r}{\lambda_F}\frac{s_F}{w_F-s_F}\in(0,1).\end{aligned}$$

Details for Equation 8

$\forall t_R\leq t_F$,

$$\begin{aligned}W_R(t_R,t_F) =& e^{-rt_R}(p_0+(1-p_0)e^{-\lambda_R t_R})E_{t_R}\left[\int_{t_R}^{+\infty}w_Re^{-r(t-t_R)}dt-(1-\boldsymbol{p_R}(t_R))\mathbb{1}_{\tilde{t}<t_F}\int_{\tilde{t}}^{+\infty}c_Re^{-r(t-t_R)}dt\right]\\ =& e^{-rt_R}\frac{p_0}{\boldsymbol{p_R}(t_R)}\left[\frac{w_R}{r}-(1-\boldsymbol{p_R}(t_R))\frac{c_R}{r}e^{(r+\lambda_F)t_R}\int_{t_R}^{t_F}\lambda_Fe^{-(r+\lambda_F)t}dt\right]\\ =& e^{-rt_R}\frac{p_0}{\boldsymbol{p_R}(t_R)}\left[\frac{w_R}{r}-(1-\boldsymbol{p_R}(t_R))\frac{c_R}{r}\frac{\lambda_F}{r+\lambda_F}(1-e^{-(r+\lambda_F)(t_F-t_R)})\right].\end{aligned}$$

As for the refuter's payoff, the researcher's payoff can be expressed as a function of p_R and p_F, which is, for all $p_R\leq p_F^*$,

$$\begin{aligned}\tilde{W}_R(p_R,p_F) &= W_R\left(\frac{1}{\lambda_R}\left(\log\big(\frac{1-p_0}{p_0}\big)+\log\big(\frac{p_R}{1-p_R}\big)\right),\frac{1}{\lambda_R}\log\big(\frac{1-p_0}{p_0}\big)+\frac{\lambda_R-\lambda_F}{\lambda_F\lambda_R}\log\big(\frac{1-p_R}{p_R}\big)+\frac{1}{\lambda_F}\log\big(\frac{p_F}{1-p_F}\big)\right)\\ &= \frac{p_0}{p_R}\left(\frac{p_0}{1-p_0}\frac{1-p_R}{p_R}\right)^{\frac{r}{\lambda_R}}\left[\frac{w_R}{r}-(1-p_R)\frac{c_R}{r}\frac{\lambda_F}{r+\lambda_F}\left(1-\left(\frac{1-p_F}{p_F}\frac{p_R}{1-p_R}\right)^{\frac{r+\lambda_F}{\lambda_F}}\right)\right].\end{aligned}$$

Proof of Proposition 2

Define, for all $p_R < p_F^*$,

$$\tilde{\tilde{W}}_R(p_R) = \left(\frac{p_0}{1-p_0}\frac{1-p_R}{p_R}\right)^{\frac{r}{\lambda_R}}\frac{p_0}{p_R}\left[\frac{w_R}{r} - (1-p_R)\frac{c_R}{r}\frac{\lambda_F}{r+\lambda_F}\left(1-\left(\frac{1-p_F^*}{p_F^*}\frac{p_R}{1-p_R}\right)^{\frac{r+\lambda_F}{\lambda_F}}\right)\right].$$

By doing the change of variable $X = \dfrac{1-p_R}{p_R}$,

$$V_R(X) = \tilde{\tilde{W}}_R(m(X)) = \tilde{\tilde{W}}_R(p_R) \text{ with } m(X) = \frac{1}{1+X}.$$

$m(.)$ is a decreasing function of X from values p_F^* to 0 with $\lim_{X\to+\infty} m(X) = 0$.

$\forall X \in \left[X_F^*, +\infty\right)$, with $X_F^* = \dfrac{1-p_F^*}{p_F^*}$,

$$\begin{aligned}
V_R(X) =& p_0\left(\frac{p_0}{1-p_0}\right)^{\frac{r}{\lambda_R}} X^{\frac{r}{\lambda_R}}\left[(1+X)\frac{w_R}{r} - X\frac{\lambda_F}{r+\lambda_F}\frac{c_R}{r} + \frac{\lambda_F}{r+\lambda_F}\frac{c_R}{r}\left(X_F^*\right)^{\frac{r+\lambda_F}{\lambda_F}} X^{-\frac{r}{\lambda_F}}\right],\\
V_R'(X) =& p_0\left(\frac{p_0}{1-p_0}\right)^{\frac{r}{\lambda_R}} X^{\frac{r}{\lambda_R}-1}\frac{1}{r\lambda_R}\left[rw_R - X(r+\lambda_R)\left(\frac{\lambda_F}{r+\lambda_F}c_R - w_R\right)\right.\\
&\left. - (\lambda_R-\lambda_F)\frac{r}{r+\lambda_F}c_R\left(X_F^*\right)^{\frac{r+\lambda_F}{\lambda_F}} X^{-\frac{r}{\lambda_F}}\right].
\end{aligned}$$

Define, $\forall X \in \left[0, +\infty\right)$,

$$\begin{aligned}
f(X) &= rw_R - X(r+\lambda_R)\left(\frac{\lambda_F}{r+\lambda_F}c_R - w_R\right) - (\lambda_R-\lambda_F)\frac{r}{r+\lambda_F}c_R(X_F^*)^{\frac{r+\lambda_F}{\lambda_F}} X^{-\frac{r}{\lambda_F}},\\
f'(X) &= -(r+\lambda_R)\left(\frac{\lambda_F}{r+\lambda_F}c_R - w_R\right) + (\lambda_R-\lambda_F)\frac{r}{\lambda_F}\frac{r}{r+\lambda_F}c_R(X_F^*)^{\frac{r+\lambda_F}{\lambda_F}} X^{-\frac{r+\lambda_F}{\lambda_F}},\\
f''(X) &= -(\lambda_R-\lambda_F)\frac{r^2}{\lambda_F^2}c_R(X_F^*)^{\frac{r+\lambda_F}{\lambda_F}} X^{-\frac{r+\lambda_F}{\lambda_F}-1}.
\end{aligned}$$

There are two cases: either $\lambda_R > \lambda_F$ or $\lambda_R \leq \lambda_F$.

1. $\lambda_R > \lambda_F$: f'' is negative for all X, hence, f is concave. Moreover,

$$\lim_{X\to 0} f(X) = -\infty,$$
$$\lim_{X\to +\infty} f(X) = -\infty.$$

   For small values of X, f' is positive and for large values of X, f' is negative. Hence, f is increasing then decreasing on $\big[0,+\infty\big)$, taking possibly positive values.

2. $\lambda_R \leq \lambda_F$: f'' is positive and f' negative for all X, hence, f is decreasing and convex. Moreover,

$$\lim_{X\to 0} f(X) = +\infty,$$
$$\lim_{X\to +\infty} f(X) = -\infty.$$

   Hence, for small values of X, f is positive and for large values of X, f is negative.

In both cases, the equilibrium researcher's choice depends on the sign of $f\Big(X_F^*\Big)$.

$$\begin{aligned} f\Big(X_F^*\Big) &= rw_R + (r+\lambda_R)X_F^* w_R - \lambda_R X_F^* c_R \\ &= \frac{1}{p_F^*}\Big(rw_R + \lambda_R(1-p_F^*)(w_R - c_R)\Big), \\ f(X_F^*) = 0 \Leftrightarrow c_R &= \left(1 + \frac{\lambda_F}{\lambda_R}\frac{w_F - s_F}{s_F}\right) w_R. \end{aligned}$$

Define $\bar{c}_R \equiv \left(1 + \frac{\lambda_F}{\lambda_R}\frac{w_F - s_F}{s_F}\right) w_R$. The different cases are:

1. If $c_R \geq \bar{c}_R$, i.e. $f(X_F^*) < 0$, then for every X higher than X_F^*, f is always negative. V_R is decreasing on $[X_F^*, +\infty)$. In other words, W_R is increasing on $(0, p_F^*]$.

(a) If $p_0 < p_F^*$, $\max_{p_0 \le p_R \le p_F^*} \tilde{\tilde{W}}_R(p_R) = \tilde{\tilde{W}}_R(p_F^*)$ hence, $p_R^* = p_F^*$;

(b) Otherwise $\max_{p_0 \le p_R \le p_F^*} \tilde{\tilde{W}}_R(p_R) = \tilde{\tilde{W}}_R(p_0) = \tilde{\tilde{W}}_R(p_F^*)$ and the researcher publishes immediately at time 0, i.e. $p_R^* = p_0$;

2. If $c_R < \bar{c}_R$, i.e. $f(X_F^*) > 0$, then there exists a $X_R^* > X_F^*$ such that $f(X_R^*) = 0$; f is positive on $[X_F^*, X_R^*)$ and negative on $[X_R^*, +\infty)$. V_R increases on $[X_F^*, X_R^*)$ and decreases on $[X_R^*, +\infty)$. In other words, $\tilde{\tilde{W}}_R$ is increasing on $(0, m(X_R^*)]$ and decreasing on $(m(X_R^*), p_F^*]$.

(a) If $p_0 < \bar{p}_0$ with $\bar{p}_0 = m(X_R^*)$, $\max_{0 \le p_R \le p_F^*} \tilde{\tilde{W}}_R(p_R) = \tilde{\tilde{W}}_R(m(X_R^*))$, hence, $p_R^* = m(X_R^*)$;

(b) Otherwise $\max_{p_0 \le p_R \le p_F^*} \tilde{\tilde{W}}_R(p_{R,t_R}) = \tilde{\tilde{W}}_R(p_0)$ and the researcher publishes at time 0, i.e. $p_R^* = p_0$.

Details of Figure 2

The equilibrium switching quality p_F^* does not depend on the refutation cost c_R and the refutation cost's threshold $\bar{c}_R$ does not depend on p_0, the initial belief. However, the threshold cost $\bar{c}_R$ depends on the initial belief through the equality $\tilde{\tilde{W}}'(\bar{p}_0) = 0$. It is easier to analyze the cost in terms of the initial belief, which is:

$$\begin{aligned}
\tilde{\tilde{W}}'(\bar{p}_0) = 0 \Leftrightarrow & f\left(\frac{1-\bar{p}_0}{\bar{p}_0}\right) = 0 \\
\Leftrightarrow & c_R(\bar{p}_0) = \frac{w_R(r+\lambda_F)(r+\lambda_R+r\frac{\bar{p}_0}{1-\bar{p}_0})}{\lambda_F(r+\lambda_R) + r(\lambda_R-\lambda_F)\left(\frac{1-p_F^*}{p_F^*}\frac{\bar{p}_0}{1-\bar{p}_0}\right)^{\frac{r+\lambda_F}{\lambda_F}}}.
\end{aligned}$$

Besides,

$$c_R(0) = \frac{r+\lambda_F}{\lambda_F} w_R,$$
$$c_R(p_F^*) = \bar{c}_R.$$

1. If $\lambda_R \leq \lambda_F$, it is obvious that c_R is increasing in $\overline{p}_0$.

2. If $\lambda_R > \lambda_F$, then the monotonicity of c_R has to be proven.

$\forall \overline{p}_0 \in [0, p_F^*]$,

$$c'_R(\overline{p}_0) = \frac{r(r+\lambda_F)w_R\Big[\lambda_F^2(r+\lambda_R) - (\lambda_R-\lambda_F)\big(\frac{1-p_F^*}{p_F^*}\frac{\overline{p}_0}{1-\overline{p}_0}\big)^{\frac{r+\lambda_F}{\lambda_F}}\big(\lambda_F(r+\lambda_R) + r(\lambda_R + \frac{r}{1-\overline{p}_0})\big)\Big]}{\lambda_F(1-\overline{p}_0)^2\Big(\lambda_F(r+\lambda_R) + r(\lambda_R-\lambda_F)\big(\frac{1-p_F^*}{p_F^*}\frac{\overline{p}_0}{1-\overline{p}_0}\big)^{\frac{r+\lambda_F}{\lambda_F}}\Big)^2}.$$

Denote, for all $\overline{p}_0$ in $[0, p_F^*]$,

$$g(\overline{p}_0) = \lambda_F^2(r+\lambda_R) - (\lambda_R-\lambda_F)\Big(\frac{1-p_F^*}{p_F^*}\frac{\overline{p}_0}{1-\overline{p}_0}\Big)^{\frac{r+\lambda_F}{\lambda_F}}\Big(\lambda_F(r+\lambda_R) + r(\lambda_R + \frac{r}{1-\overline{p}_0})\Big).$$

$g(.)$ is decreasing in $\overline{p}_0$. Hence, its minimum is achieved at p_F^*.

$$\begin{aligned} g(p_F^*) &= \lambda_F^2(r+\lambda_R) - (\lambda_R-\lambda_F)\Big(\lambda_F(r+\lambda_R) + r(\lambda_R + \lambda_F\frac{w_F-s_F}{s_F})\Big), \\ \frac{\partial g(p_F^*)}{\partial \lambda_R} &= \lambda_F^2 - \Big(\lambda_F(r+\lambda_R) + r(\lambda_R+\lambda_F\frac{w_F-s_F}{s_F})\Big) - (\lambda_R-\lambda_F)(r+\lambda_F) \\ &= \lambda_F(2\lambda_F - r\frac{w_F-s_F}{s_F}) - 2\lambda_R(r+\lambda_F) < 0. \end{aligned}$$

Hence, $g(p_F^*)$ decreases in λ_R. Moreover,

$$\lim_{\lambda_R \to +\infty} g(p_F^*) = -\infty,$$
$$\lim_{\lambda_R \to \lambda_F} g(p_F^*) = \lambda_F^2(r + \lambda_F) > 0.$$

Consequently, by the intermediate value theorem, there exists a $\overline{\lambda}_R$ in $(\lambda_F, +\infty)$ such that $g(p_F^*)$ is positive for λ_R in $(\lambda_F, \overline{\lambda}_R]$ and negative for λ_R in $(\overline{\lambda}_R, +\infty)$. Hence, for λ_R in $(\lambda_F, \overline{\lambda}_R]$, $c_R'(\overline{p}_0)$ is positive for all $\overline{p}_0$ in $[0, p_F^*]$, meaning that $c_R(.)$ is increasing in $\overline{p}_0$ on $[0, p_F^*]$. Otherwise, for λ_R in $(\overline{\lambda}_R, +\infty)$, $c_R'(\overline{p}_0)$ is positive then negative for $\overline{p}_0$ in $[0, p_F^*]$, meaning that $c_R(.)$ is increasing then decreasing on $[0, p_F^*]$. On Figure 2 is represented the case where λ_R belongs to $(\lambda_F, \overline{\lambda}_R]$.

Proof of Lemma 1

In case of pessimistic risky publication, the refutation cost has an upper bound, which is,

$$c_R < \left(1 + \frac{\lambda_F}{\lambda_R}\frac{w_F - s_F}{s_F}\right) w_R.$$

As this refutation cost must satisfy Inequality 3 in Assumption 1, which is $c_R > \dfrac{r + \lambda_F}{\lambda_F} w_R$, a necessary condition for the existence of this case is:

$$\frac{r + \lambda_F}{\lambda_F} < 1 + \frac{\lambda_F}{\lambda_R}\frac{w_F - s_F}{s_F} \Leftrightarrow \lambda_F > \sqrt{\frac{r\lambda_R s_R}{w_F - s_F}}.$$

Proof of Proposition 3

The derivative of f with respect to c_R, evaluated at X_R^* is:

$$\left.\frac{\partial f}{\partial c_R}\right|_{X=X_R^*} = -X_R^*(r+\lambda_R)\frac{\lambda_F}{r+\lambda_F} - (\lambda_R-\lambda_F)\frac{r}{r+\lambda_F}\left(X_F^*\right)^{\dfrac{r+\lambda_F}{\lambda_F}}\left(X_R^*\right)^{-\dfrac{r}{\lambda_F}}.$$

1. If $\lambda_R > \lambda_F$, then it is negative. Hence, $f(X_R^*)$ and $V_R'(X_R^*)$ decrease in c_R. X_R^* decreases in c_R. As a consequence, p_R^* increases in c_R.

2. If $\lambda_R \leq \lambda_F$, then it is decreasing in X_R^*. Hence, it achieves its maximum in X_F^*.

$$\left.\frac{\partial f}{\partial c_R}\right|_{X=X_R^*} < -\lambda_R X_F^* < 0$$

Hence, $\left.\frac{\partial f}{\partial c_R}\right|_{X=X_R^*}$ is always negative. p_R^* increases in c_R.

For comparative statics with respect to w_R, I compute:

$$\left.\frac{\partial f}{\partial w_R}\right|_{X=X_R^*} = r + X_F^*(r+\lambda_R) > 0.$$

Hence, $f(X_R^*)$ and $V_R'(X_R^*)$ increase in w_R. X_R^* increases in w_R. Consequently, p_R^* decreases in w_R.

For comparative statics with respect to λ_R, I compute:

$$\left.\frac{\partial f}{\partial \lambda_R}\right|_{X=X_R^*} = -X_R^*\left(\frac{\lambda_F}{r+\lambda_F}c_R - w_R\right) - \frac{r}{r+\lambda_F}c_R\left(X_F^*\right)^{\dfrac{r+\lambda_F}{\lambda_F}}\left(X_R^*\right)^{-\dfrac{r}{\lambda_F}} < 0.$$

$f(X_R^*)$ and $V_R'(X_R^*)$ decrease in λ_R so X_R^* decreases in λ_R. Hence, p_R^* increases in λ_R.

For comparative statics with respect to λ_F, I compute:

$$\left.\frac{\partial f}{\partial \lambda_F}\right|_{X=X_R^*} = \frac{rc_R}{\lambda_F^2(r+\lambda_F)^2}\left(X_R^*\right)^{-\frac{r}{\lambda_F}}\left(X_F^*\right)^{\frac{r+\lambda_F}{\lambda_F}}\Bigg[-\lambda_F^2(r+\lambda_R)\big(X_R^*/X_F^*\big)^{\frac{r+\lambda_F}{\lambda_F}} - (\lambda_R-\lambda_F)r(r+\lambda_F)\log\left(X_R^*/X_F^*\right)$$
$$- \frac{\lambda_F}{\lambda_F w_F - (r+\lambda_F)s_F}\big((\lambda_F^3 - \lambda_R r^2 + \lambda_F^2(r-2\lambda_R) + r\lambda_F(r-2\lambda_R))w_F - (r+\lambda_F)(\lambda_F^2 - 2\lambda_F\lambda_R - r\lambda_R)s_F\big)\Bigg].$$

$p_R^* < p_F^*$ hence $X_F^* < X_R^*$ hence $\log\left(X_R^*/X_F^*\right) > 0$. Define, for $X \in [X_F^*, +\infty)$,

$$\begin{aligned}
h(X) =& -\lambda_F^2(r+\lambda_R)\big(X/X_F^*\big)^{\frac{r+\lambda_F}{\lambda_F}} - (\lambda_R-\lambda_F)r(r+\lambda_F)\log\left(X/X_F^*\right)\\
&- \frac{\lambda_F}{\lambda_F w_F - (r+\lambda_F)s_F}\big((\lambda_F^3 - \lambda_R r^2 + \lambda_F^2(r-2\lambda_R) + r\lambda_F(r-2\lambda_R))w_F - (r+\lambda_F)(\lambda_F^2 - 2\lambda_F\lambda_R - r\lambda_R)s_F\big),\\
h'(X) =& -\lambda_F(r+\lambda_R)(r+\lambda_F)\big(X\big)^{\frac{r}{\lambda_F}}\left(X_F^*\right)^{\frac{r+\lambda_F}{\lambda_F}} + (\lambda_F-\lambda_R)r(r+\lambda_F)\big(X\big)^{-1}\\
=& X^{-1}(r+\lambda_F)\Big(-\lambda_F(r+\lambda_R)\big(X/X_F^*\big)^{\frac{r+\lambda_F}{\lambda_F}} + r(\lambda_F-\lambda_R)\Big).
\end{aligned}$$

h' is clearly decreasing in X, hence, it achieves its maximum at X_F^*. For all X in $[X_F^*, +\infty)$,

$$\begin{aligned}
h'(X) &\le h'(X_F^*)\\
&\le -\left(X_F^*\right)^{-1}(r+\lambda_F)^2\lambda_R\\
&\le 0.
\end{aligned}$$

Hence, h decreases in X and achieves its maximum at X_F^*.

- When $\lambda_R \le \lambda_F$,

$$\begin{aligned}
\left.\frac{\partial f}{\partial \lambda_F}\right|_{X=X_R^*} &< \left.\frac{\partial f}{\partial \lambda_F}\right|_{X=X_F^*} = -(\lambda_F-\lambda_R)\frac{r^2 s_F c_R(w_F-s_F)}{\lambda_F(\lambda_F w_F - (r+\lambda_F)s_F)^2}\\
&< 0.
\end{aligned}$$

Hence, $f(X_R^*)$ and $V'(X_R^*)$ decrease in λ_F. X_R^* decreases in λ_F. Consequently, p_R^*

increases in λ_F when $\lambda_R \leq \lambda_F$.

- When $\lambda_R > \lambda_F$, as X_R^* decreases in λ_R, it achieves its lower bound at λ_F. Denote $\underline{X}_R^*$ this lower bound. It is defined as

$$\lim_{\lambda_R \to \lambda_F} f(\underline{X}_R^*) = 0 \Leftrightarrow \underline{X}_R^* = \frac{r w_R}{\lambda_F c_R - (r + \lambda_F) w_R}.$$

For any X_R^*,

$$\begin{aligned} h(X_R^*) &\leq h(\underline{X}_R^*) \\ &\leq \lambda_F^2 (r + \lambda_F) \left[1 - \left(\underline{X}_R^* / X_F^* \right)^{\dfrac{r + \lambda_F}{\lambda_F}} \right]. \end{aligned}$$

In the pessimistic risky case, $c_R < \overline{c_R}$. When λ_R tends to λ_F, it is equivalent to $c_R / w_R < w_F / s_F$, which ensures that $\underline{X}_R^* > X_F^*$. Hence, the right part of the inequality is negative, which implies that $h(X_R^*)$ is always negative. In this way, $f(X_R^*)$ and $V'(X_R^*)$ decrease in λ_F. X_R^* decreases in λ_F. Consequently, p_R^* increases in λ_F when $\lambda_R > \lambda_F$.

For comparative statics with respect to w_F, I compute:

$$\left. \frac{\partial f}{\partial w_F} \right|_{X = X_R^*} = (\lambda_R - \lambda_F) c_R \frac{r^2 s_F}{(\lambda_F (w_F - s_F) - r s_F)^2} \left(X_F^* \right)^{\dfrac{r}{\lambda_F}} \left(X_R^* \right)^{-\dfrac{r}{\lambda_F}},$$

which is positive if $\lambda_R > \lambda_F$. In this case, $f(X_R^*)$ and $V_R'(X_R^*)$ increase in w_F, hence, X_R^* increases in w_F and p_R^* decreases in w_F. Otherwise, which is if $\lambda_R \leq \lambda_F$, p_R^* increases in w_F.

For comparative statics with respect to s_F, I compute:

$$\left.\frac{\partial f}{\partial s_F}\right|_{X=X_R^*} = -(\lambda_R - \lambda_F)c_R \frac{r^2 w_F}{(\lambda_F(w_F - s_F) - r s_F)^2}\left(X_F^*\right)^{\dfrac{r}{\lambda_F}}\left(X_R^*\right)^{-\dfrac{r}{\lambda_F}},$$

which is negative if $\lambda_R > \lambda_F$. In this case, $f(X_R^*)$ and $V_R'(X_R^*)$ decrease in s_F, hence, X_R^* decreases in s_F and p_R^* increases in s_F. Otherwise, which is if $\lambda_R \leq \lambda_F$, p_R^* decreases in s_F.

Proof of Lemma 2

The participation constraint of an informed type is

$$W_R^i(t_R^i) \geq 0,$$

where t_R^i is the publication time of an informed researcher and

$$\begin{aligned} W_R^i(t_R^i) &= e^{-rt_R^i}(1-p_0)(1-e^{-\lambda_R t_R^i})E_{t_R^i}\left[\int_{t_R^i}^{+\infty} w_R e^{-r(t-t_R^i)}dt - \int_{\tilde{t}}^{+\infty} c_R e^{-r(t-t_R^i)}dt\right] \\ &= e^{-rt_R^i}(1-p_0)(1-e^{-\lambda_R t_R^i})\left[\frac{w_R}{r} - \frac{c_R}{r}\frac{\lambda_F}{r+\lambda_F}\right] < 0. \end{aligned}$$

This participation constraint is never satisfied under Assumption 1. Hence, there exists no separating equlibrium in pure strategies.

Proof of Proposition 4

The refuter's payoff is:

$$\begin{aligned} W_F^P(t_R, t_F) =& E_0\left[p_0\int_{t_F}^{+\infty} s_F e^{-rt}dt + (1-p_0)\left(\mathbb{1}_{\tilde{t}<t_F}\int_{\tilde{t}}^{+\infty} w_F e^{-rt}dt + \mathbb{1}_{\tilde{t}>t_F}\int_{t_F}^{+\infty} s_F e^{-rt}dt\right)\right] \\ =& \frac{1}{r}e^{-rt_F}\left[p_0 s_F + (1-p_0)s_F e^{-\lambda_F(t_F-t_R)} + \frac{\lambda_F}{r+\lambda_F}(1-p_0)w_F\left(e^{r(t_F-t_R)} - e^{-\lambda_F(t_F-t_R)}\right)\right]. \end{aligned}$$

Taking the first order condition with respect to t_F, I obtain the refuter's equilibrium experimentation time Δt^{**}:

$$\begin{aligned}
\Delta t^{**} &= \max\{0; \frac{1}{\lambda_F}\log\Big(\frac{1-p_0}{p_0}\frac{\lambda_F w_F-(r+\lambda_F)s_F}{rs_F}\Big)\} \\
&= \begin{cases} \frac{1}{\lambda_F}\log\Big(\frac{1-p_0}{p_0}\frac{\lambda_F w_F-(r+\lambda_F)s_F}{rs_F}\Big) & \text{if } p_0 < 1-\frac{r}{\lambda_F}\frac{s_F}{w_F-s_F} \\ 0 & \text{otherwise.} \end{cases}
\end{aligned}$$

In the latter case, i.e. if $p_0 \geq 1-\frac{r}{\lambda_F}\frac{s_F}{w_F-s_F}$, the refuter never experiments and switches directly to the safe option.

Proof of Proposition 5

$$\begin{aligned}
W_R^i(t_R) &= e^{-rt_R}(1-p_0)(1-e^{-\lambda_R t_R})E_{t_R}\left[\int_{t_R}^{+\infty} w_R e^{-r(x-t_R)}dx - \mathbb{1}_{\tilde{t}<t_R+\Delta t^{**}}\int_{\tilde{t}}^{+\infty} c_R e^{-r(x-t_R)}dx\right] \\
&= e^{-rt_R}(1-p_0)(1-e^{-\lambda_R t_R})\left[\frac{w_R}{r} - \frac{c_R}{r}\frac{\lambda_F}{r+\lambda_F}\left(1-\left(\frac{1-p_F^*}{p_F^*}\frac{p_0}{1-p_0}\right)^{\frac{r+\lambda_F}{\lambda_F}}\right)\right], \\
W_R^i(t) &= e^{-rt}(1-p_0)(1-e^{-\lambda_R t})E_t\left[\int_t^{+\infty} w_R e^{-r(x-t)}dx - \int_{\tilde{t}}^{+\infty} c_R e^{-r(x-t)}dx\right] \\
&= e^{-rt}(1-p_0)(1-e^{-\lambda_R t})\left[\frac{w_R}{r} - \frac{c_R}{r}\frac{\lambda_F}{r+\lambda_F}\right] < 0. \\
W_R^u(t_R) &= e^{-rt_R}\big(p_0+(1-p_0)e^{-\lambda_R t_R}\big)E_{t_R}\left[\int_{t_R}^{+\infty} w_R e^{-r(x-t_R)}dx - (1-\boldsymbol{p_R}(t_R))\mathbb{1}_{\tilde{t}<t_R+\Delta t^{**}}\int_{\tilde{t}}^{+\infty} c_R e^{-r(x-t_R)}dx\right] \\
&= e^{-rt_R}\big(p_0+(1-p_0)e^{-\lambda_R t_R}\big)\left[\frac{w_R}{r} - (1-\boldsymbol{p_R}(t_R))\frac{c_R}{r}\frac{\lambda_F}{r+\lambda_F}\left(1-\left(\frac{1-p_F^*}{p_F^*}\frac{p_0}{1-p_0}\right)^{\frac{r+\lambda_F}{\lambda_F}}\right)\right] \\
&= e^{-rt_R}\left[p_0\frac{w_R}{r} + (1-p_0)e^{-\lambda_R t_R}\left(\frac{w_R}{r} - \frac{c_R}{r}\frac{\lambda_F}{r+\lambda_F}\left(1-\left(\frac{1-p_F^*}{p_F^*}\frac{p_0}{1-p_0}\right)^{\frac{r+\lambda_F}{\lambda_F}}\right)\right)\right], \\
W_R^u(t) &= e^{-rt}\big(p_0+(1-p_0)e^{-\lambda_R t}\big)E_t\left[\int_t^{+\infty} w_R e^{-r(x-t)}dx - (1-\boldsymbol{p_R}(t))\int_{\tilde{t}}^{+\infty} c_R e^{-r(x-t)}dx\right] \\
&= e^{-rt}\big(p_0+(1-p_0)e^{-\lambda_R t}\big)\left[\frac{w_R}{r} - (1-\boldsymbol{p_R}(t))\frac{c_R}{r}\frac{\lambda_F}{r+\lambda_F}\right].
\end{aligned}$$

Due to Inequality (3), $W_R^i(t)$ is always negative. Hence, the incentive constraint of an informed type is always satisfied. Define $\hat{c}_R$ as

$$\hat{c}_R = \frac{r+\lambda_F}{\lambda_F}\left[1-\left(\frac{p_0}{1-p_0}\frac{rs_F}{\lambda_F w_F-(r+\lambda_F)s_F}\right)^{\frac{r+\lambda_F}{\lambda_F}}\right]^{-1} w_R.$$

If $c_R > \hat{c}_R$, the participation constraint of an informed type is never satisfied. Consequently, there exists no pooling equilibrium. If $c_R \leq \hat{c}_R$, the participation constraint of an informed type is satisfied for any $t_R \in \mathbb{R}^+$. The participation and incentive constraints for an uninformed type are always satisfied for $c_R \leq \hat{c}_R$. Hence, a pooling equilibrium exists for any $t_R \in \mathbb{R}^+$. Besides, the payoff for an uninformed researcher decreases in t_R. Hence, the pooling equilibrium which maximizes the uninformed researcher's payoff is

$t_R = 0$.

Details of Figure 3

The threshold p_F^* does not depend on the refutation cost c_R. From Equation 12, $\hat{c}_R$ is clearly increasing in p_0. Moreover, $\hat{c}_R$ is equal to $\frac{r+\lambda_F}{\lambda_F} w_R$ when the initial belief is 0 and tends to infinity when the initial belief tends to the switching quality.

Bibliography

Andrews, Isaiah and Maximilian Kasy (2019) "Identification of and correction for publication bias," *American Economic Review*, 109 (8), 2766–94.

Begley, C. Glenn and Lee M. Ellis (2012) "Drug development: Raise standards for preclinical cancer research," *Nature*, 483 (7391), 531–533.

Bobtcheff, Catherine, Jérôme Bolte, and Thomas Mariotti (2017) "Researcher's dilemma," *The Review of Economic Studies*, 84 (3), 969–1014.

Bobtcheff, Catherine and Raphaël Levy (2017) "More haste, less speed? Signaling through investment timing," *American Economic Journal: Microeconomics*, 9 (3), 148–86.

Brodeur, Abel, Mathias Le, Marc Sangnier, and Yanos Zylberberg (2016) "Star wars: The empirics strike back," *American Economic Journal: Applied Economics*, 8 (1), 1–32.

Davis, Courtney, Huseyin Naci, Evrim Gurpinar, Elita Poplavska, Ashlyn Pinto, and Ajay Aggarwal (2017) "Availability of evidence of benefits on overall survival and quality of life of cancer drugs approved by European Medicines Agency: retrospective cohort study of drug approvals 2009-13," *The British Medical Journal*, 359, j4530.

Di Tillio, Alfredo, Marco Ottaviani, and Peter Norman Sorensen (2017) "Persuasion bias in science: can economics help?" *The Economic Journal*, 127 (605), F266–F304.

Dur, Robert and Otto H. Swank (2005) "Producing and manipulating information," *The Economic Journal*, 115 (500), 185–199.

Felgenhauer, Mike and Petra Loerke (2017) "Bayesian persuasion with private experimentation," *International Economic Review*, 58 (3), 829–856.

Freedman, Leonard P., Iain M. Cockburn, and Timothy S. Simcoe (2015) "The economics of reproducibility in preclinical research," *PLoS biology*, 13 (6), e1002165.

Furukawa, Chishio (2017) "Unbiased publication bias: Theory and evidence," Unpublished Manuscript.

Hamermesh, Daniel S. (2007) "Replication in economics," *Canadian Journal of Economics*, 40 (3), 715–733.

Henry, Emeric and Marco Ottaviani (2019) "Research and the approval process: the organization of persuasion," *American Economic Review*, 109 (3), 911–55.

Hoppe-Wewetzer, Heidrun C, Georgios Katsenos, and Emre Ozdenoren (2019) "Experimentation, Learning, and Preemption," *Available at SSRN 3321231*.

Humphreys, Macartan (2015) "What has been learned from the deworming replications: a nonpartisan view," *Columbia University, Aug. http://www. columbia. edu/˜ mh2245/w/worms. html*.

Ioannidis, John P.A. (2005a) "Why most published research findings are false," *PLoS medicine*, 2 (8), e124.

——— (2005b) "Contradicted and initially stronger effects in highly cited clinical research," *Jama*, 294 (2), 218–228.

Keller, Godfrey, Sven Rady, and Martin Cripps (2005) "Strategic experimentation with exponential bandits," *Econometrica*, 73 (1), 39–68.

Kiri, Bralind, Nicola Lacetera, and Lorenzo Zirulia (2018) "Above a swamp: A theory of high-quality scientific production," *Research Policy*, 47 (5), 827–839.

Landis, Story C., Susan G. Amara, Khusru Asadullah, Chris P. Austin, Robi Blumenstein, Eileen W. Bradley, Ronald G. Crystal, Robert B. Darnell, Robert J. Ferrante, Howard Fillit et al. (2012) "A call for transparent reporting to optimize the predictive value of preclinical research," *Nature*, 490 (7419), 187–191.

Libgober, Jonathan (2018) "False Positives and Transparency in Scientific Research," *Available at SSRN 2617130*.

Maniadis, Zacharias, Fabio Tufano, and John A. List (2014) "One swallow doesn't make a summer: New evidence on anchoring effects," *American Economic Review*, 104 (1), 277–90.

McClellan, Andrew (2017) "Experimentation and Approval Mechanisms," working paper.

Moscarini, Giuseppe and Lones Smith (2001) "The optimal level of experimentation," *Econometrica*, 69 (6), 1629–1644.

Mueller-Langer, Frank, Benedikt Fecher, Dietmar Harhoff, and Gert G. Wagner (2019) "Replication studies in economics—How many and which papers are chosen for replication, and why?" *Research Policy*, 48 (1), 62–83.

Open Science Collaboration (2015) "Estimating the reproducibility of psychological science," *Science*, 349 (6251), aac4716.

Petersen, Alexander Michael, Santo Fortunato, Raj K. Pan, Kimmo Kaski, Orion Penner, Armando Rungi, Massimo Riccaboni, H. Eugene Stanley, and Fabio Pammolli (2014) "Reputation and impact in academic careers," *Proceedings of the National Academy of Sciences*, 111 (43), 15316–15321.

Salop, Steven C. (1979) "Strategic entry deterrence," *The American Economic Review*, 69 (2), 335–338.

Smaldino, Paul E. and Richard McElreath (2016) "The natural selection of bad science," *Royal Society Open Science*, 3 (9), 160384.

Stephan, Paula E. (1996) "The economics of science," *Journal of Economic literature*, 34 (3), 1199–1235.

The Economist (2013) "Unreliable research: Trouble at the lab," *Economist*.

Yoder, Nathan (2018) "Designing incentives for heterogeneous researchers," *Available at SSRN 3154143*.

Chapter 2

Citizen science vs. traditional science: the speed/quality trade-off

Abstract I examine the tradeoff of the scientist who has access to citizen science (CS) or traditional science (TS) to undertake her research. A sequence of projects are successively available on a discrete time horizon. Each project is a public good with some value. For each project, a scientist (she) chooses to implement it with CS or TS. With TS, she implements the project for sure but it takes one period. With CS, the project is implemented immediately with the help of two citizens. However, its success is uncertain and depends on the citizens' actions. When the successive projects have the same value, the scientist's strategy is a cut-off: when the value of the project is above some threshold, she uses CS. Below this threshold, she uses TS. This result is generalized to any sequence of project values which satisfies the Markov property. In an extension, two scientists compete to attract citizens on their project. Focusing on stationnary Markovian equilibria, I show that the equilibrium in pure strategies is unique: for sufficiently high project values, both scientists always choose CS. For sufficiently low values, both scientists always choose TS. In both cases, there exist no equilibrium in mixed strategies. For intermediate project values, the unique equilibrium is in mixed strategies: if one scientist is stuck with choosing TS at the former period, the other one always chooses CS. Otherwise, they both mix between TS and CS.

Keywords: citizen science, repeated game, tradeoff of the scientist.

JEL codes: C73, 031, 036.

1 Introduction

There is growing interest in citizen science, a form of science to be distinguished from traditional science. Also called "crowd science" or "crowdsourcing", it democratizes science by involving citizens to implement various kinds of projects, as Franzoni and Sauermann (2014) analyze. As reported by Koeze and Popper (2020), the covid-19 crisis has emphasized the importance of connectivity, which plays a key role in citizen science. For instance, the FoldIt project, launched in 2008, enables citizens to contribute to a science project through an online game in which they fold proteins to understand their structure. In 2010, it already gathered 200 000 players. Since last March, the game has proposed to look for the protein which will neutralize the covid-19. In another project called Galaxy Zoo, launched in 2007, citizens treat galaxies data in order to report dust particules, a task which can not be done even by highly performing computers. The success of this project led to the creation of a online platform in which volunteers have the choice to contribute to various kinds of citizen science projects. In that way, citizen science contributes largely to the collection of important data. In this respect, Poisson et al. (2020) shows that the contribution of the citizens provided over half of the observations for commonly sampled water-quality measures worldwide from the past 31 years, and also contributed to the majority of long-term monitoring for selected measures in lakes. Moreover, as stated by Lukyanenko et al. (2016), "because citizens generally lack formal scientific training, they view problems and issues in light of their own knowledge and interests, creating fertile ground for discoveries".

Besides, citizen science is a powerful tool to raise awareness on major scientific concerns. Kullenberg and Kasperowski (2016) indicates that there are three citizen science fields of research. The first and largest is research on biology, conservation and ecology, and utilizes citizen science mainly as a methodology of collecting and classifying data. Taking the example of migratory birds and climate change, Cooper et al. (2014) find that papers based on citizen science constituted between 24% and 77% of the references backing each claim, with no evidence of a mistrust of claims that relied heavily on citizen-science data. Concerning biodiversity research, Theobald et al. (2015) estimate that 1.3–2.3 million volunteers contribute $667 million–$2.5 billion in-kind annually. This gives a tool to understand the crucial impact of citizen science on public expenses. The second field is geographic information research, where citizens participate in the collection of geographic data. The third one is research on social sciences and epidemiology, which studies and facilitates public participation in relation to environmental issues and health. For example, Mahajan et al. (2020) shows that using it through online quizz can generate insightful data which can assist in understanding people's perception and exposure levels to air pollution.

Nevertheless, the quality of science made using citizen science is questionable as it involves non-experts, as underlined by Lukyanenko et al. (2016). It should be considered carefully. Indeed, Prestopnik et al. (2017) underline the importance of games to motivate participants, but cheating should be controlled for. In short, crowd science can be very useful to achieve projects in a surprinsingly short amount of time. Numerous citizen science projects share the same features: 1) a high number of volunteers participate to each project; 2) it enables scientists to implement projects which would have taken years otherwise; 3) each project is a public good which is privately implemented by the volunteers and 4) the quality of the project's outcomes is questionable. As reported by Rowland (2012), Francois Taddei, a molecular geneticist at the Centre for Research and Interdisciplinarity at Paris Descartes University, suggests that a solution to the worrying quality of citizen science projects may be to develop a "science of citizen science". More specifically, researchers will need to figure out the optimal division of labour between citizens and professionals.

To meet these goals, I build a theoretical model to better understand the mechanisms behind citizen science. There exists an infinite number of scientific projects successively available. Each project is a public good, which means that every player enjoys its value, which is called "the project value". There are three players, one scientist and two citizens. For each project, the scientist (she) has the choice between using citizen science (CS) or traditional science (TS) to implement it. By using TS, she works by herself on the project and does not require the help of the citizens. To implement the project, she needs one period of time and the next project is available only two periods after. By using CS, the scientist makes the citizens work to implement the project at the current period, but its success depends on the actions of the citizens. The next project is available only one period after. To use CS, the scientist must pay a fixed cost, which is called "the CS cost". This cost represents the cost of making the project available to citizens online. The creation of a platform through which citizens can contribute always represents a sunk cost for the scientist. Taking the example of FoldIt, designing the game through which citizens contribute to the project is costly. The CS cost can also represent the cost of sharing data. Indeed, these data are sometimes hard to obtain and making it public is a sunk cost for the scientist.

The citizens only work if the scientist has chosen CS for the current period's project. Each one chooses to exert effort or not. The effort cost is privately observed by each citizen and depends on his motivation. If he is motivated, then effort is costless. If he is not motivated, then effort is costly. The probability of implementing the project depends on how many citizens exerted effort. If no one exerts effort, then the project is not implemented. If only one exerts effort, then it is implemented with some probability. If both exert effort, it is implemented with a higher probability. In any case, the quality of the project is lower with CS. To simplify the analysis, I assume that the non-motivated citizen never exerts effort, which implies that there exists only one citizens' Bayesian Nash equilibrium.

In a first part, I analyze the deterministic benchmark, in which each project has the same value. By backward induction, I first look at the Bayesian Nash equilibrium of the citizens. Due to the assumption discussed above, the equilibrium is unique: a motivated citizen always exerts effort and a non-motivated one never exerts effort. The ex-ante probability that the project is implemented is called "the CS quality". Turning to the scientist, her payoff is linear in the scientist's choice and does not involve a state variable. It implies that her choice is constant and follows a cut-off rule. If the project value is above the cut-off, the scientist chooses CS and if it is below, she chooses TS. This cut-off decreases with the CS quality and increases with the CS cost and the discount factor, which is quite intuitive. Indeed, if the CS quality decreases, the discount factor increases or the CS cost increases, TS becomes an attractive option and the scientist chooses TS for higher project values.

Then, I generalize the deterministic benchmark by applying the model to any sequence of project values which satisfies the Markov property. Using recursive techniques (Stokey and Lucas, 1989), I show that there always exists a cut-off for the project value, above which the scientist uses CS and below which she uses TS to implement the project.

In an extension, instead of one scientist, two scientists compete to attract citizens. By contrast with the monopolist case, projects are not public goods: one scientist does not benefit from the implementation of the project of the other scientist. The probability to attract citizens is a modified Tullock function with a minimal CS cost to pay. This function is well suited to my model as it increases with the CS cost of one scientist and decreases with the CS cost of the other scientist. By contrast with the monopolist case, the cost represents the effort that one scientist should make to attract citizens and cannot be recovered for the loser of the contest. The benefit coming from the implementation of the project is random. By backward induction, the strategies of the citizens remain unchanged. Then looking at the strategies of the scientists, I restrain the analysis to Markovian equilibria, that is, equilibria in which at every period, each scientist chooses always the same CS cost. The Markovian state variable is the scientist's current project value. It means that scientists cannot adopt repeated strategies like punishment phases. It implies that both scientists choose the same CS cost every time they both choose CS. Besides, they have equal probability to attract the two citizens on their project. The "simplest" Markovian equilibrium to look at is the stationary equilibrium, in which each scientist makes the same choice at every period.

In this stationary Markovian equilibrium, a scientist is "constrained" if she has chosen TS for the former period's project, meaning that she has no choice to make at the current period. A scientist is "not constrained" if she has chosen CS at the former period, meaning that a new project is available for her at the current period. The analysis starts by fixing the strategy of one scientist and looking at a profitable deviation for the other one. The latter has two types of choice to make. The first type occurs when the other scientist is constrained, which implies that she is a monopolist to attract citizens. The second type of choice occurs when the other one is not constrained, which implies that they compete to attract citizens.

There exist three kinds of outcomes. In the first one, both scientists choose TS. In the two remaining outcomes, both scientists choose CS, with the equilibrium CS being equal to the minimum in one case and a higher one in the second case. The cut-off project value to use TS is the same as the monopolist case as the continuation value from choosing TS remains the same. The cut-off to use CS is similar but scientists compete so the CS quality is multiplied by one half. Besides, the monopolist CS cost is replaced by the equilibrium CS cost of the static Nash game played by scientists. Interestingly, there exists no equilibrium in which a scientist chooses CS if both are not constrained and TS if only the other one is constrained. Also, there exists no equilibrium in which one chooses TS if both are not constrained and CS if only the other one is constrained. This comes from the fact that choosing CS in one case implies that the project value must be sufficiently high and at the same time, choosing TS in the other case implies that it must be sufficiently low. These two conditions cannot hold at the same time for these two kinds of equilibria. By contrast with the monopolist case, there exist two sources of inefficiencies. The first one is that when both scientists always choose CS and the equilibrium CS cost is higher than the minimum, they pay an extra CS cost. The second one is that the chance to win the contest is only a half, implying that project value threshold to choose CS is higher. In that way, there exists no equilibrium in pure strategies when the project value lies between the TS threshold and the CS threshold.

Turning to mixed strategies, a unique symmetric equilibrium exists only in the range of project values where no equilibrium in pure strategies exists. In this equilibrium, a scientist always chooses CS when the other one is constrained and scientists mix between TS and CS when both are not constrained. Similarly as the analysis of equilibria in pure strategies, no other type of equilibrium in mixed strategies exists. To conclude the analysis, two sources of randomness appear in this mixed strategies equilibrium. The first one comes from the winner-takes-all contest to attract citizens. The loser pays the sunk CS cost and does not implement her project at all. The second one comes from mixing between CS and TS when both scientists are not constrained. The benefit of randomly picking CS when the other scientist picks TS implies that the scientist pays the minimal CS cost over two periods, thus enjoying the monopolist position longer. In this way, a persistence effect is engendered.

Related Literature This chapter is related to the private provision of public good field, which Bergstrom et al. (1986) explore in detail in their seminal paper. They study a model in which the citizens contribute in a non-cooperative way to the provision of a public good. They show that there exists in general a unique Nash equilibrium. In my model, in addition, each citizen does not know the other one's type. Consequently, I look for a Bayesian Nash equilibrium. Bagnoli and Lipman (1989) show with an experiment that citizens always have sufficient incentives to voluntarily achieve the Pareto efficient outcome. In this chapter, I introduce a probability of providing the public good which depends on the effort exerted by both citizens. In this way, the Pareto efficient outcome, which would be that both citizens exert effort, is not implemented. Moreover, my modeling of citizens' game is similar to Varian (2004), which compares three different technologies for the probability of implementing the project. In my model, I rather keep a general modeling of this probability to focus on the scientist's tradeoff.

Besides, my model is in line with the research process literature, an economic field which focuses on the scientist's work. Bramoullé and Saint-Paul (2010) builds a model of scientific progress in which scientists decide to work on an old paradigm or to create a new one. Furthermore, Garfagnini and Strulovici (2012) emphasize the trade-off between marginal innovation and radical innovation to measure the intergenerational accumulation of knowledge. By contrast, in my model, the scientist trades off between traditional science and citizen science to work on a new project. In addition, I focus on the probability of implementing a project rather than on the accumulation of knowledge created by all the projects implemented.

Finally, this chapter contributes to the citizen science literature. In this respect, Watson and Floridi (2016) emphasizes the powerful epistemic advantages of citizen science over more traditional modes of scientific investigation. I model these advantages with the researcher going faster by choosing citizen science. Besides, Nov et al. (2014) investigate effects of motivational factors on the quantity and quality of citizen scientists' contribution. They show that citizens' contributions are determined by collective motives, norm-oriented motives, reputation, and intrinsic motives, and that contribution quality is positively affected only by collective motives and reputation. In my model, I suppose that the quality of the citizen science project depends on the quantity of contribution made by the citizens. Interestingly, Laut et al. (2017) underline that citizen science projects often rely on only a small fraction of participants to make the majority of contributions. To illustrate this fact, in my model, I suppose that there are only two citizens. They analyze the impact of including virtual peers to citizen science projects on the citizens contribution. They found that when virtual peers systematically outperform the participants, their contribution increases fourfold. In my model, when the two citizens are motivated, the quality of the outcome is higher.

A close paper to this chapter is Bernard (2020). This paper also analyzes the researchers' tradeoff between CS and TS. At the time at which one of the researchers discovers a new idea/project, she chooses between CS, which brings a certain payoff, and TS. By contrast with this chapter, the researcher who uses TS chooses the time at which she publishes her idea. The more the idea mature, the higher her payoff. However, she faces potential competition from her opponent as they fight for priority. In this way, Bernard (2020) studies a model in which scientists compete when using CS but not TS and this chapter studies a model in which they compete using TS but not CS.

The chapter is organized as follows. Section II explains the model. In Section III, the deterministic benchmark is analyzed and in Section IV generalizes the result of Section III. In Section V, an extension is provided where 2 scientists compete to attract citizens. Section VI concludes.

2 The Model

Time is denoted $t \in \mathbb{N}^*$ and discounted at rate $\delta \in (0, 1)$. There exists an infinite number of scientific projects, which are successively available. Each project is a public good, so that if the n^{th} project is implemented, each player enjoys its value $v_n > 0$. If the project is not implemented, each player gets no payoff. There exist three players, one scientist and two citizens.

2.1 The scientist

If a project is available at date t, the scientist (she) has the choice between using citizen science (CS) or traditional science (TS) to implement it. If she uses TS, she works alone on the project and the latter is implemented at date $t + 1$. The next project can only be available at date $t + 2$. If she chooses CS, she requires the citizens' help to complete the project. The latter is implemented at the current period, that is, at date t, but there is some probability that the project fails, depending on the citizens' actions. In this case, every player gets 0 and the project can never be implemented. Besides, after having observed the current project value, the scientist must pay a cost $c_{inf} > 0$ to make the project available to citizens. This cost is called *the CS cost*. By using CS, the next project is available at date $t + 1$. I assume that if the scientist is indifferent between CS and TS for one project, then she always chooses CS.

2.2 The citizens

If the scientist chooses CS for the project available at date t, then citizens work. A citizen (he) may be motivated with probability $\alpha \in (0,1)$, or not (with probability $1-\alpha$). Each citizen $j \in \{1,2\}$ chooses to exert effort (E) or not (NE). Exerting effort costs $e > 0$ for a non-motivated citizen and 0 for a motivated one. Each citizen faces a double uncertainty: he does not know whether the other one is motivated and if the latter exerts effort or not. The probability of implementing the project, P_m, depends on how many citizens exert effort. If both exert effort, the project is implemented with probability $P_2 \in (0,1]$. If only one exerts effort, it is implemented with probability $P_1 \in (0, P_2)$. If none exerts effort, the project is implemented with probability $P_0 = 0$.

Assumption 1.

$$\forall n \in \mathbb{N}^*,\ e > (P_2 - P_1)v_n. \tag{1}$$

Inequality 1 means that a non-motivated citizen never exerts effort for the n^{th} project, even if the other does.

3 The deterministic benchmark

Each project has the same value $v > 0$. By backward induction, I start by studying the strategies of the citizens.

3.1 The citizens' Bayesian Nash equilibrium

I look for a Bayesian Nash equilibrium. Denote e_{-j} the effort cost of citizen $-j$. If citizen j is motivated, he faces the following game:

j $-j$	E	NE
E	$P_2 v$, $P_2 v - e_{-j}$	$P_1 v$, $P_1 v$
NE	$P_1 v$, $P_1 v - e_{-j}$	0 , 0

As effort is costless for a motivated citizen, it is always a dominant strategy for him to exert effort. If citizen j is not motivated, he faces the following game:

j $-j$	E	NE
E	$P_2 v - e$, $P_2 v - e_{-j}$	$P_1 v - e$, $P_1 v$
NE	$P_1 v$, $P_1 v - e_{-j}$	0 , 0

Due to Inequality 1, it is always a dominant strategy for a non-motivated citizen not to exert effort. The analysis remains the same for citizen $-j$. This leads to the following lemma.

Lemma 1. *The Bayesian Nash equilibrium is unique: a motivated citizen always exerts effort and a non-motivated one never exerts effort.*

This Lemma implies that the ex-ante probability that the project is implemented is:

$$\mathbb{P}(\text{project is implemented}) = \mathbb{E}[P_m] = \alpha^2 P_2 + 2\alpha(1-\alpha)P_1 = f_\alpha. \quad (2)$$

In what follows, f_α is called *the CS quality.* The CS quality is always lower than TS quality, that always equals 1 (the project never fails).

3.2 The tradeoff of the scientist

By choosing CS, the scientist's expected payoff at date t is $f_\alpha v - c_{inf}$. By choosing TS, her expected payoff at date t is δv. The scientist's choice for the project available at date t is denoted $k_t \in \{0, 1\}$. $k_t = 0$ means that she chooses TS and $k_t = 1$ that she chooses CS for this project. The scientist's value function is the following:

$$V = \max_{k_t \in \{0,1\}} \Big\{ k_t \Big[f_\alpha v - c_{inf} + \delta V \Big] + (1-k_t) \Big[\delta v + \delta^2 V \Big] \Big\};$$

as there is no state variable, the scientist's choice is stationary and equal to $k \in \{0,1\}$:

$$= \max_{k \in \{0,1\}} \Big\{ k \Big[f_\alpha v - c_{inf} + \delta V \Big] + (1-k) \Big[\delta v + \delta^2 V \Big] \Big\}$$

$$= \max_{k \in \{0,1\}} \Big\{ k \Big[(f_\alpha - \delta) v - c_{inf} + \delta(1-\delta) V \Big] \Big\} + \delta v + \delta^2 V$$

The function to maximize is linear in k. For each k, the value function is linear in v. In this way, there exists a unique threshold v^* for the project value defined as follows.

$$v^*(f_\alpha) \equiv \begin{cases} \dfrac{c_{inf}}{f_\alpha - \dfrac{\delta}{1+\delta}} & \text{if } f_\alpha > \dfrac{\delta}{1+\delta}; \\ +\infty & \text{otherwise.} \end{cases}$$

Proposition 1. *The scientist uses CS for every project if $v \geq v^*(f_\alpha)$. Otherwise, she uses TS.*

Since CS involves a sunk cost for the scientist, she chooses this option only if the project value is high enough. If the CS quality is below $\dfrac{\delta}{1+\delta}$, the value function when choosing CS is always lower than when choosing TS, which implies that CS is always more costly than TS to implement the project. As a consequence, the scientist always chooses TS. Otherwise, the scientist may choose CS, if the project value is above $v^*(f_\alpha)$. The threshold $v^*(f_\alpha)$ increases with the CS cost in citizen science and the discount factor and decreases with the CS quality. Indeed, when c_{inf} or δ is high or when f_α is low, the citizen science's choice becomes poorly attractive and the scientist requires higher project values. The equilibrium choice of the scientist is represented in Figure 1. The following section generalizes the cut-off rule of the scientist to any sequence of projects which satisfies the Markov property.

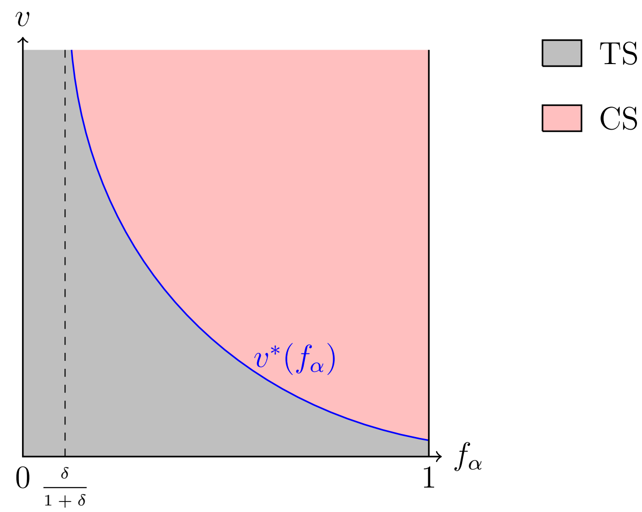

Figure 1 – The equilibrium choice of the scientist on the graph (f_α, v)

4 Generalization

Suppose that every project has a value $v_n \in [\underline{v}, \bar{v}]$ with $\bar{v} > \underline{v} \geq 0$. Denote v the current project value and v^+ the value of the next project, which is not yet available. Suppose that the evolution of the project values satisfies the Markov property, that is, the value of the next project only depends on the last project value through its distribution $G(v^+|v)$, and that this distribution is increasing in v in the FOSD sense, that is:

$$\forall v \geq v', \forall v^+, G(v^+|v) \leq G(v^+|v'). \tag{3}$$

It means that if the scientist has a low (respectively high) value for the currently available project, she has more chances to have a low (respectively high) value for the next project available. In this respect, there exists a persistency effect between the project which is currently available and the one which is available just after.

The citizens' strategy remains unchanged: a motivated citizen always exerts effort and a non-motivated one never exerts effort. Accordingly, the CS quality remains the same as Equation 2. The scientist's value function becomes:

$$\begin{aligned} TV(v) &= \max_{k\in[0,1]} \Big\{ k\Big[f_\alpha v - c_{inf} + \delta\mathbb{E}[V(v^+)|v]\Big] + (1-k)\Big[\delta v + \delta^2\mathbb{E}[V(v^+)|v]\Big]\Big\} \\ &= \max_{k\in[0,1]} \Big\{ \delta v + k\Big((f_\alpha - \delta)v - c_{inf}\Big) + \delta(k+\delta(1-k))\int_{\underline{v}}^{\bar{v}} V(v^+)g(v^+|v)dv^+ \Big\}. \end{aligned}$$

Inequality 3 implies that the scientist making a choice at time t has a high probability to make the same choice for the next project available. In this way, either she chooses CS or TS for the two projects. If the current project value is high, then she chooses CS, and most probably CS for the next project. Otherwise, she chooses TS, and most probably TS for the next project. By applying the contraction mapping theorem to the scientist's dynamic problem, the result found in Section 3 can be generalized.

Proposition 2. *There exists a unique threshold $v^{**} \in [\underline{v}, \bar{v}]$ such that, for any project available with value v:*

- *If $v \geq v^{**}$, the scientist chooses CS;*
- *Otherwise, she chooses TS.*

In this way, the uncertainty about the next projects outcomes is controlled by this cut-off rule. This simple strategy is easy to implement for the scientist which has only to care about the current project value to make her choice. In addition, this result enables to make a research plan over several periods. In terms of policy implications, it is important as the government can anticipate the potential subsidies to provide to support CS.

4.1 Application to the iid case

This framework is a special case of the Markovian evolution of the projects values. The sequence $(v_n)_{n\in\mathbb{N}^*}$ is independent and identically distributed on $\left[\underline{v}, \bar{v}\right]$ with distribution

F and mean E_v and $\bar{v} > \underline{v} > 0$. To compare with Inequality 3, the distribution F satisfies

$$\forall v \geq v', \forall v^+, F(v^+|v) = F(v^+|v') = F(v^+).$$

In this respect, if the currently available project has a high value, the next project has not necessarily a high value. No matter if the current project value is high or low, the expected continuation payoff remains the same. The next project value plays no role as the latter is not correlated with the currently available project value. This makes the scientist's choice easy as she only takes into account the current project value to make her decision. Besides, there exists no persistence in the scientist's strategy. Choosing CS for the current project does not imply a high probability of choosing CS for the next project. In terms of policy implications, this framework does not necessarily benefit the government (he). Indeed, he cannot accurately plan how to provide funds to encourage CS for future research projects.

Up to now, the scientist was certain to capt the two citizens attentions on her project using CS. What happens if two scientists fight for the citizens? This is what I study in the following Section.

5 When scientists compete to attract citizens

The evolution of the project values is deterministic and equal to $v > 0$. By contrast with the monopolist case, projects are not public goods, in the sense that one scientist cannot benefit from the implementation of the other scientist's project. Scientists fight to attract citizens and the loser gets 0. Scientist i has a probability $p(c_i, c_{-i})$ to attract citizens on her project, which depends on her CS investment c_i and on the other scientist's CS investment c_{-i}. In this sense, the CS investment differs from the monopolist case in which the scientist is certain to attract citizens. This probability is a modified Tullock function with a minimal CS investment c_{inf} to pay, that is,

$$p(c_i, c_{-i}) = \begin{cases} 1 & \text{if } c_i \geq c_{inf} \text{ and } c_{-i} < c_{inf}; \\ \dfrac{c_i}{c_i + c_{-i}} & \text{if } c_i \geq c_{inf} \text{ and } c_{-i} \geq c_{inf}; \\ 0 & \text{otherwise.} \end{cases}$$

This probability is increasing in the CS investment of scientist i and decreasing in the CS investment of scientist $-i$. Observe that the CS investment cannot be recovered for the loser of the contest. To compare with the monopolist scientist case, $c_i - c_{inf}$ is a loss created entirely by the contest. The benefit, $f_\alpha v$, which represents the implementation of the project with CS, is random. If both scientists choose the same CS investment, they have the same probability to win the contest. In short, the expected benefit with CS under competition is $p(c_i, c_{-i}) f_\alpha v$.

Using backward induction, the strategies of the citizens remain the same, that is, a motivated citizen always exerts effort and a non-motivated one never exerts effort. Turning to the strategies of the scientists, I focus on Markovian equilibria, that is, equilibria in which at every period, each scientist chooses always chooses the same CS cosst. The Markovian state variable is the scientist's current project value. By choosing CS, the expected payoff of scientist i at date t is $p(c_i, c_{-i}) f_\alpha v - c_i$. By choosing TS, her expected payoff at date t is δv. If both scientists choose CS, they play a pure strategies Nash equilibrium, with scientist i choosing $c_i^{BR}(c_{-i})$ such that:

$$\begin{aligned} c_i^{BR}(c_{-i}) &= \arg \max_{c_i \in [c_{inf}, +\infty)} \{p(c_i, c_{-i}) f_\alpha v - c_i\} \\ &= \begin{cases} \sqrt{f_\alpha v c_{-i}} - c_{-i} & \text{if } \sqrt{f_\alpha v c_{-i}} - c_{-i} \geq c_{inf}; \\ c_{inf} & \text{otherwise.} \end{cases} \end{aligned}$$

Since the game is symmetric, they choose the same equilibrium CS investment.

Lemma 2. *If the scientists play the same strategies at every period, they both choose the equilibrium CS investment c^*, equal to*

$$c^* \equiv \max\{\frac{f_\alpha v}{4}, c_{inf}\}.$$

As a consequence, each scientist has the same probability $p(c^*, c^*) = 1/2$ to attract citizens. If $c_{inf} > \frac{f_\alpha v}{4}$, they invest the minimal CS investment c_{inf} and enjoy at the current period $\frac{1}{2} f_\alpha v - c_{inf}$. Otherwise, they invest $\frac{f_\alpha v}{4}$ and enjoy at the current period $\frac{f_\alpha v}{4}$. In the following Section, I look for a stationary Markovian equilibrium.

5.1 The stationnary Markovian equilibrium

A scientist is *constrained* if she has chosen TS for the former period's project, meaning that she has no choice to make at the current period. By contrast, a scientist is *not constrained* if she has chosen CS at the former period, meaning that a new project is available at the current period. Denote V_{11} the continuation value of scientist i if both scientists are not constrained, V_{10} the continuation value if scientist i is not constrained but scientist $-i$ is, V_{01} the value function if scientist i is constrained but not scientist $-i$ and V_{00} the value function if both scientists are constrained. Denote k_{11} the equilibrium choice of the scientists if they are both constrained and k_{10} the equilibrium choice of scientist i if scientist $-i$ is constrained. Fix scientist $-i$'s strategy to k_{11} and k_{10} and look at a profitable deviation for scientist i. Her continuation values and value functions are equal to

$$
\begin{aligned}
V_{11} =& \max_{k\in[0,1]}\{kk_{11}(\frac{1}{2}f_\alpha v - c^* + \delta V_{11}) + k(1-k_{11})(f_\alpha v - c_{inf} + \delta V_{10}) + (1-k)k_{11}\delta V_{01} + (1-k)(1-k_{11})\delta V_{00}\};\\
V_{10} =& \max_{k\in[0,1]}\{k(f_\alpha v - c_{inf} + \delta V_{11}) + (1-k)\delta V_{01}\};\\
V_{01} =& v + \delta[k_{10}V_{11} + (1-k_{10})V_{10}];\\
V_{00} =& v + \delta V_{11}.
\end{aligned}
$$

If both scientists are not constrained, scientist i's value function V_{11} splits in four terms. First, with probability kk_{11}, both scientists choose CS and they compete to attract citizens. At the next period, they are not constrained and enjoy the continuation payoff V_{11}. Second, with probability $k(1-k_{11})$, scientist i chooses CS but not scientist $-i$ and she only has to pay the minimal CS investment c_{inf} to catch citizens. At the next period, scientist $-i$ is constrained but not scientist i and she enjoys the continuation payoff V_{10}. Third, with probability $(1-k)k_{11}$, scientist i chooses TS and scientist $-i$ CS. The former is constrained at the next period and enjoys the continuation value V_{01}. Fourth, with probability $(1-k)(1-k_{11})$, both scientists choose TS and they are constrained at the next period. Scientist i enjoys the continuation value V_{00} at the next period.

If scientist $-i$ is constrained but not scientist i, scientist i's value function V_{10} splits in 2 terms. First, with probability k, she chooses CS and enjoys the monopolist payoff. At the next period, both scientists are not constrained and she enjoys the continuation payoff V_{11}. Second, with probability $1-k$, she chooses TS. At the next period, she is constrained but not scientist $-i$ and enjoys the continuation value V_{01}.

If scientist i is constrained but not scientist $-i$, then she has no choice to make at the current period and she enjoys the payoff v. If scientist $-i$ has chosen CS at the current period, then at the next period she is not constrained and scientist i enjoys the continuation value V_{11}. If scientist $-i$ has chosen TS at the current period, then at the next period, she is constrained but not scientist i and the latter enjoys the continuation payoff V_{10}.

If both scientists are constrained, scientist i enjoys the payoff v and at the next period, both scientists are not constrained and scientist i enjoys the continuation payoff V_{11}.

Replacing V_{01} and V_{00} into V_{11} and V_{10}, I obtain

$$V_{11} = \max_{k\in[0,1]}\{k\Big[k_{11}((\frac{1}{2}f_\alpha - \delta)v - c^* + \delta(1-\delta k_{10})V_{11} - \delta^2(1-k_{10})V_{10})$$
$$+(1-k_{11})((f_\alpha - \delta)v - c_{inf} + \delta V_{10} - \delta^2 V_{11})\Big] + k_{11}\delta[v + \delta(k_{10}V_{11} + (1-k_{10})V_{10})] + (1-k_{11})\delta(v + \delta V_{11})\};$$
$$V_{10} = \max_{k\in[0,1]}\{k[(f_\alpha - \delta)v - c_{inf} + \delta(1-\delta k_{10})V_{11} - \delta^2(1-k_{10})V_{10}] + \delta[v + \delta(k_{10}V_{11} + (1-k_{10})V_{10})]\}.$$

which are linear in k. In the following section, I focus on symmetric equilibria in pure strategies in which the scientists have the same strategies.

Symmetric equilibria in pure strategies

There exist three kinds of equilibrium outcomes. Either both scientists choose TS or both scientists choose CS with an equilibrium CS investment equal to c_{inf}, or both choose CS with an equilibrium CS investment equal to $\frac{f_\alpha v}{4}$. Suppose that when the project value is below the monopolist cut-off $v^*(f_\alpha)$, the equilibrium is such that both scientists choose TS. Then no scientist has an interest in deviating by choosing CS as she would not choose it without a competitor. Consequently, this equilibrium exists. Above the cutoff $v^*(f_\alpha)$, a scientist have an interest in deviating from this type of equibrium as she would obtain the CS monopolist payoff, which is higher than the TS payoff. The cutoff to choose CS is similar to $v^*(f_\alpha)$ and equal to

$$\frac{c^*}{\frac{f_\alpha}{2} - \frac{\delta}{1+\delta}}.$$

With competition, each scientist only enjoys half of the attention of the citizens, that is why f_α is multiplied by 1/2. Besides, each pays the CS investment c^*, instead of the monopolist case, which was always c_{inf}. When the CS investment is $\frac{f_\alpha v}{4}$, the scientist's current payoff with CS becomes $\frac{f_\alpha v}{4}$. In this way, the CS quality must be at least four times higher than the monopolist case, which was $\frac{\delta}{1+\delta}$. In addition, if the discount factor is too high, then the TS option becomes too attractive so that scientists deviates from the CS equilibrium. Denote $\tilde{f}_\alpha$, $\tilde{v}_1(f_\alpha)$, $\tilde{v}_2(f_\alpha)$ and such that

$$\begin{aligned}
\tilde{f}_\alpha &\equiv \frac{4\delta}{1+\delta}; \\
\tilde{v}_1(f_\alpha) &\equiv \frac{4c_{inf}}{f_\alpha}; \\
\tilde{v}_2(f_\alpha) &\equiv \frac{c_{inf}}{\frac{f_\alpha}{2} - \frac{\delta}{1+\delta}}.
\end{aligned}$$

Proposition 3. *The symmetric equilibrium in pure strategies is unique:*

- *If* $f_\alpha \geq \tilde{f}_\alpha$, $v \geq \tilde{v}_1(f_\alpha)$ *and* $\delta < \frac{1}{3}$, *then both scientists choose CS and* $c^* = \frac{f_\alpha v}{4}$*;*
- *If* $\tilde{v}_2(f_\alpha) \leq v < \tilde{v}_1(f_\alpha)$ *and* $\delta < \frac{1}{3}$, *then both scientists choose CS and* $c^* = c_{inf}$*;*
- *If* $v < v^*(f_\alpha)$, *then both scientists always choose TS.*

If the CS quality is lower than $\tilde{f}_\alpha$, then using CS is too costly for scientists, who have to compete and only enjoy half of the expected CS quality. If the project value is above $\tilde{v}_1(f_\alpha)$, then CS becomes very attractive and scientists are ready to pay a high

CS investment, $\frac{f_\alpha v}{4}$, to catch citizens. If the project value is between $\tilde{v}_2(f_\alpha)$ and $\tilde{v}_1(f_\alpha)$, then both scientist choose CS and each pays c_{inf} to attract half of the citizens.

Interestingly, there exists no equilibrium in which $k_{11} = 1$ and $k_{10} = 0$. Indeed, the expected benefit from choosing CS is always higher when the other scientist is constrained. Hence, it cannot be that the scientist chooses CS when both are not constrained and TS when the other is constrained. There exists no equilibrium in which $k_{11} = 0$ and $k_{10} = 1$. To choose CS when the other scientist is constrained, it should be that the project value is sufficiently high. To choose TS when both scientists are not constrained, it should be that the project value is sufficiently low. These two conditions cannot hold at the same time.

The three outcomes are represented in Figure 2 if the discount factor is below $\tilde{\delta}$. When this latter is above the threshold, the violet and pink areas do not appear on the graph. To compare with the deterministic monopolist case, when the project value lies below $v^*(f_\alpha)$, competition has no impact on the scientist's strategy. The latter always chooses TS and does not involve citizens. However, when the project value is above $v^*(f_\alpha)$, there exist two sources of inefficiencies. The first one is that in the violet area, scientists pay a CS investment higher than the minimum c_{inf} to catch citizens. This loss is created by the contest game. The second inefficiency is that the CS choice, represented in violet and pink in Figure 2, requires a higher CS quality. It comes from the fact that each scientist only has one chance out of two to win the contest against the other. Consequently, there exists no equilibrium in pure strategies in the white area. In the following Section, I look for the existence of a mixed strategies equilibrium in these area.

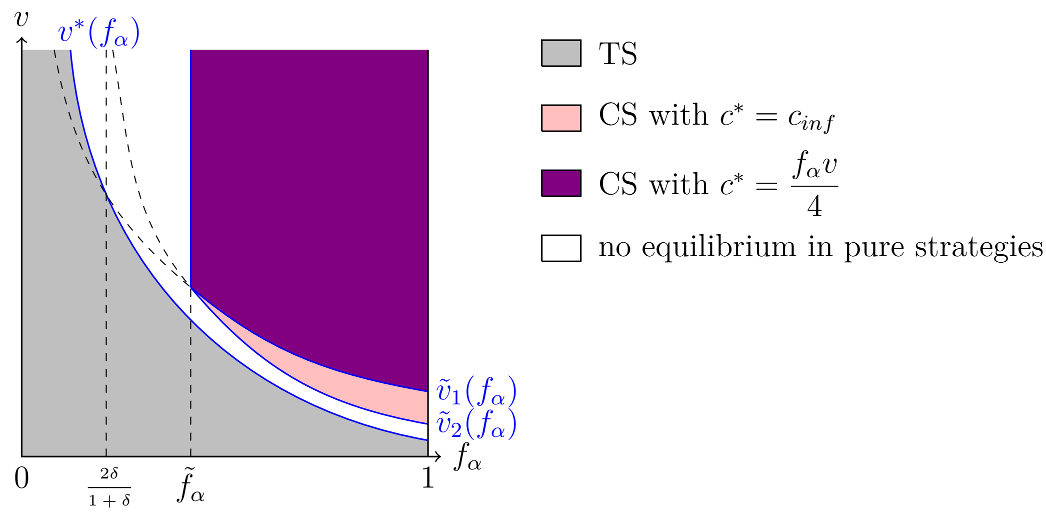

Figure 2 – Symmetric equilibria in pure strategies on the graph (f_α, v) if $\delta < \frac{1}{3}$

Symmetric equilibria in mixed strategies

Intuitively, competition decreases the expected payoff of one scientist. Hence, the value function when scientists are not constrained is always lower than when only the competitor is constrained. In this way, if $k_{10} = 0$, then $k_{11} = 0$. This removes equilibria of the type: i) $k_{11} = 1$ and $k_{10} \in (0, 1)$ and ii) $k_{11} \in (0, 1)$ and $k_{10} = 0$. Besides, no equilibrium of the type $k_{11} = 0$ and $k_{10} \in (0, 1)$ exists for the same reason as no equibrium of the type $k_{11} = 0$ and $k_{10} = 1$ exists. If both scientists are not constrained and mix between TS and CS, then the TS option is attractive. It means that the project value is low enough. If the competitor is constrained and the scientist mixes between CS and TS, then the CS option is attractive. It means that the project value is high enough. As a matter of fact, these two conditions cannot be true at the same time. Consequently, there exists no equilibrium with $k_{11} \in (0, 1)$ and $k_{10} \in (0, 1)$. The only existing symmetric equilibrium in mixed strategies is the following.

Proposition 4. *If $\{f_\alpha^* < f_\alpha < \tilde{f}_\alpha$ and $v \geq v^*(f_\alpha)\}$ or $\{f_\alpha > \tilde{f}_\alpha$ and $v^*(f_\alpha) \leq v <$*

$\tilde{v}_2(f_\alpha)\}$, there exists a unique symmetric equilibrium in mixed strategies characterized by the following scientists' strategy:

- *If both scientists are not constrained, they mix between CS and TS. If they both pick CS,*
  - *If $v < \tilde{v}_1$, they pay $c^* = \dfrac{f_\alpha v}{4}$;*
  - *Otherwise, they pay $c^* = c_{inf}$;*
- *If one scientist is constrained, then the other always chooses CS.*

Together with the pure strategies equilibria, this equilibrium forms a partition of the graph (f_α, v). Interestingly, two sources of randomness appear in this equilibrium. The first one comes from the contest between the scientists to attract citizens. The winner attracts the two citizens on her project and the loser pays the sunk CS investment, without enjoying any benefit. At the next period, both scientists remain not constrained. Hence, this source of randomness has no implication for what happens next.

The second one comes from the fact that, when both scientists are not constrained, they mix between TS and CS. It induces a persistence effect over two periods: when one scientist randomly picks TS at equilibrium, her competitor benefits from her monopoly situation. Over two periods, she only pays the minimum CS investment, c_{inf} to catch citizens, when the other scientist cannot work on the next project. At the current period, this situation induces no loss of efficiency as both projects are implemented and the CS investment is minimized. Besides, there is no learning by doing here: the unlucky scientist is stuck in a non-optimal equilibrium which only benefits her opponent. This can be analyzed as a lock-in effect.

If both scientists randomly pick CS at equilibrium, then they compete to attract citizens. In this case, only one project is implemented and the other one is throwned away. This creates a loss of efficiency for the current period. If both scientists randomly picks TS, then there is no use of CS and no contest. Both projects are implemented with certainty but it requires time. By contrast, the monopolist scientist chooses CS for this range of project values. Accordingly, to compare with the monopolist case, there is an efficiency loss in the speed at which projects are implemented.

6 Conclusion

In this chapter, I analyzed the scientists' strategies when they trade-off between TS and CS. When the scientist is a monopolist, there exists a unique cut-off above which she always chooses CS, and below it she chooses TS. Introducing a competitor implies that in the unique symmetric Markovian equilibrium in pure strategies: i) the cut-off to choose TS is the same as the monopolist case; ii) to choose CS, the minimum CS quality must be higher than in the monopolist case; iii) for intermediate project values, there exists a unique equilibrium in mixed strategies but no equilibrium in pure strategies.

This chapter allows to question the government policy to adopt regarding CS. Should the government (he) encourage the use of CS or restrict it only to one project? The answer is not obvious if two scientists compete to attract citizens. If the projects value is low, both scientists choose TS. As a matter of fact, a restriction on the CS use has no effect on the scientists' strategy. If scientists choose CS, then more projects are implemented than with TS. However, if the projects value is high, both scientists use CS with high CS costs. Accordingly, there are two inefficiencies losses. The first one is due to the winner-takes-all contest and the second one due to high CS costs. When the scientists use CS with minimal CS costs, only the first inefficiency remains. The action of the government depends on his objective. If he is interested in implementing projects faster, then he should not restrict the use of CS. If instead, he is interested in minimizing the efficiency losses at the current period, then he should restrict the CS use only to one project.

For intermediate projects values, in the mixed strategies equilibrium, it can happen that one scientist uses CS and the competitor uses TS. In this case, the CS cost is minimized and the project implemented with TS is for sure of good quality. Consequently, there is no efficiency loss. This argument is in favor of no government intervention. But it can also happen that both scientists picks TS, which is bad as the speed of projects implementation is slowed down compared to the monopolist case. This argument encourages the use of CS with subsidies to lower CS costs. If both scientists picks CS, then the analysis remains the same as high projects values.

To go further, taking a look at the duopolist case, the inefficiencies created by the symmetric equilibrium could be avoided with an asynchronous equilibrium. Indeed, by letting the other behave as a monopolist for two periods, one scientist would choose TS. Two periods after, the roles would be reversed to avoid paying a high cost due to the contest. In this way, the scientists' projects would always be implemented with certainty and without efficiency loss. Besides, the duopolist analysis could be enriched by looking at Markovian equilibrium with punishment strategies. For instance, for sufficiently high project values, an equilibrium in which both scientists choose CS with the minimum CS cost would be sustainable. In this way, high sunk CS costs could be prevented.

7 Appendix

Proof of Proposition 1

The function to maximize is linear in k. The scientist's optimal choice k^* depends on the sign of

$$(f_\alpha - \delta)v - c_{inf} + \delta(1-\delta)V.$$

If $(f_\alpha - \delta)v - c_{inf} + \delta(1-\delta)V \geq 0$, then $k^* = 1$. In this case, the continuation value V is equal to

$$V = f_\alpha v - c_{inf} + \delta V$$
$$\Leftrightarrow V = \frac{1}{1-\delta}\Big[f_\alpha v - c_{inf}\Big].$$

Replacing V into the inequality, the scientist chooses CS for every project if

$$(f_\alpha - \delta)v - c_{inf} + \delta\Big[f_\alpha v - c_{inf}\Big] \geq 0$$
$$\Leftrightarrow \Big[(1+\delta)f_\alpha - \delta\Big]v \geq (1+\delta)c_{inf}$$

If $f_\alpha < \dfrac{\delta}{1+\delta}$, the inequality never holds. Hence, the scientist uses CS if the project value is above this cut-off:

$$v^*(f_\alpha) \equiv \begin{cases} \dfrac{c_{inf}}{f_\alpha - \dfrac{\delta}{1+\delta}} & \text{if } f_\alpha > \dfrac{\delta}{1+\delta}; \\ +\infty & \text{otherwise.} \end{cases}$$

If $(f_\alpha - \delta)v - c_{inf} + \delta(1-\delta)V < 0$, then $k^* = 0$. In this case, the continuation value

V is equal to

$$V = \delta v + \delta^2 V \Leftrightarrow V = \frac{\delta}{1-\delta^2} v.$$

Replacing V into the inequality, the scientist chooses TS for every project if

$$\begin{aligned} (f_\alpha - \delta)v - c_{inf} + \frac{\delta^2}{1+\delta} v < 0 \\ \Leftrightarrow \Big[f_\alpha - \frac{\delta}{1+\delta}\Big] v < c_{inf} \end{aligned}$$

which is the opposite of the inequality for the CS choice. Consequently, the scientist uses TS if the project value is below $v^*(f_\alpha)$.

Proof of Proposition 2

Denote $\mathcal{C}([\underline{v}, \bar{v}])$ the set of bounded continuous functions on $[\underline{v}, \bar{v}]$. If V is bounded and continuous, then the function MV, defined as $MV = \int_{\underline{v}}^{\bar{v}} V(v^+) g(v^+|v) dv^+$ for every $v \in [\underline{v}, \bar{v}]$, is bounded and continuous. If follows that the function to maximize is continuous. Moreover, the set $[0, 1]$ is compact. Hence, from the theorem of the maximum, the maximum is attained and the function TV is continuous. In this way, the operator T maps $\mathcal{C}([\underline{v}, \bar{v}])$ into $\mathcal{C}([\underline{v}, \bar{v}])$.

Take V and V' such that

$$\forall v \in [\underline{v}, \bar{v}]\ ,\ V(v) \leq V'(v).$$

Fix $v \in [\underline{v}, \bar{v}]$. Then

$$V(v) \leq V'(v) \Rightarrow E(V(v^+)|v) \leq E(V'(v^+)|v)$$
$$\Rightarrow TV(v) \leq TV'(v)$$

Hence, T is increasing in V. To apply Blackwell's sufficient conditions for a contraction, T must also verify the discounting property. Take $V \in \mathcal{C}([\underline{v}, \bar{v}])$, $a \geq 0$ and $v \in [\underline{v}, \bar{v}]$.

$$[T(V+a)](v) = \max_{k\in[0,1]} \Big\{\delta v + k\Big((f_\alpha - \delta)v - c_{inf}\Big) + \delta(k + \delta(1-k)) \int_{\underline{v}}^{\bar{v}} (V(v^+) + a) g(v^+|v) dv^+\Big\}$$
$$\leq (TV)(v) + \delta a.$$

Hence, T satisfies the Blackwell's sufficient conditions and is a contraction with modulus δ. Since $\mathcal{C}([\underline{v}, \bar{v}])$ is a Banach space, then from the contraction mapping theorem, T has a unique fixed point $V \in \mathcal{C}([\underline{v}, \bar{v}])$.

Furthermore, let $\mathcal{C}'([\underline{v}, \bar{v}]) \subset \mathcal{C}([\underline{v}, \bar{v}])$ be the set of bounded, continuous and nondecreasing functions on $[\underline{v}, \bar{v}]$, and let $\mathcal{C}''([\underline{v}, \bar{v}]) \subset \mathcal{C}([\underline{v}, \bar{v}])$ be the set of strictly increasing functions. Denote H the function:

$$H : [\underline{v}, \bar{v}] \times [0,1] \to \mathbb{R}$$
$$v \times k \mapsto \delta v + k\Big((f_\alpha - \delta)v - c_{inf}\Big) + \delta(k + \delta(1-k)) \int_{\underline{v}}^{\bar{v}} V(v^+) g(v^+|v) dv^+.$$

Fix k in $[0,1]$, v and v' in $[\underline{v}, \bar{v}]$ such that $v > v'$. Then, from Inequality 3,

$$\int_{\underline{v}}^{\bar{v}} V(v^+) g(v^+|v) dv^+ \geq \int_{\underline{v}}^{\bar{v}} V(v^+) g(v^+|v') dv^+$$
$$\Rightarrow H(v,k) > H(v',k).$$

Hence, for each k, $H(\cdot, k)$ is strictly increasing. Hence, $T[\mathcal{C}'([\underline{v}, \bar{v}])] \subseteq \mathcal{C}''([\underline{v}, \bar{v}])$. By the corollary 1 to the contraction mapping theorem in Stokey and Lucas (1989), V^* is strictly increasing in v. Hence, there exists a unique threshold $v^{**} \in [\underline{v}, \bar{v}]$ such that $k^* = 1$ for $v \geq v^{**}$ and $k^* = 0$ otherwise.

Proof of Lemma 2

$$
\begin{aligned}
c_i^{BR}(c_{-i}) &= \arg \max_{c_i \in [c_{inf}, +\infty[} \{p(c_i, c_{-i}) f_\alpha v - c_i\} \\
&= \arg \max_{c_i \in [c_{inf}, +\infty[} \{\frac{c_i}{c_i + c_{-i}} f_\alpha v - c_i\} \\
&= \begin{cases} \sqrt{f_\alpha v c_{-i}} - c_{-i} & \text{if } \sqrt{f_\alpha v c_{-i}} - c_{-i} \geq c_{inf}; \\ c_{inf} & \text{otherwise.} \end{cases}
\end{aligned}
$$

Similarly,

$$
c_{-i}^{BR}(c_i) = \begin{cases} \sqrt{f_\alpha v c_i} - c_i & \text{if } \sqrt{f_\alpha v c_i} - c_i \geq c_{inf}; \\ c_{inf} & \text{otherwise.} \end{cases}
$$

Hence, the equilibrium is symmetric and $c_i = c_{-i} = c$ such that:

$$
\begin{aligned}
& c = \sqrt{c f_\alpha v} - c \\
\Leftrightarrow\ & 4c^2 = c f_\alpha v \\
\Leftrightarrow\ & c = \frac{f_\alpha v}{4}.
\end{aligned}
$$

Then the equilibrium CS investment c^* is

$$
c^* = \max\{\frac{f_\alpha v}{4}, c_{inf}\}.
$$

Proof of Proposition 3

$$
\begin{aligned}
V_{11} = \max_{k\in[0,1]} &\{kk_{11}(\frac{1}{2}f_\alpha v - c^* + \delta V_{11}) + k(1-k_{11})(f_\alpha v - c_{inf} + \delta V_{10}) \\
&+ (1-k)k_{11}\delta[v + \delta(k_{10}V_{11} + (1-k_{10})V_{10})] + (1-k)(1-k_{11})\delta(v + \delta V_{11})\} \\
= \max_{k\in[0,1]} &\{k\Big[k_{11}((\frac{1}{2}f_\alpha - \delta)v - c^* + \delta(1-\delta k_{10})V_{11} - \delta^2(1-k_{10})V_{10}) \\
+(1-k_{11})&((f_\alpha - \delta)v - c_{inf} + \delta V_{10} - \delta^2 V_{11})\Big] + k_{11}\delta[v + \delta(k_{10}V_{11} + (1-k_{10})V_{10})] + (1-k_{11})\delta(v + \delta V_{11})\};
\end{aligned}
$$

$$
\begin{aligned}
V_{10} &= \max_{k\in[0,1]} \{k[f_\alpha v - c_{inf} + \delta V_{11}] + (1-k)\delta[v + \delta(k_{10}V_{11} + (1-k_{10})V_{10})]\} \\
&= \max_{k\in[0,1]} \{k[(f_\alpha - \delta)v - c_{inf} + \delta(1-\delta k_{10})V_{11} - \delta^2(1-k_{10})V_{10}] + \delta[v + \delta(k_{10}V_{11} + (1-k_{10})V_{10})]\}.
\end{aligned}
$$

There are 4 cases.

1. If $k_{11}\Big((\frac{1}{2}f_\alpha - \delta)v - c^* + \delta(1-\delta k_{10})V_{11} - \delta^2(1-k_{10})V_{10}\Big) + (1-k_{11})\Big((f_\alpha - \delta)v - c_{inf} + \delta V_{10} - \delta^2 V_{11}\Big) \geq 0$ and $(f_\alpha - \delta)v - c_{inf} + \delta(1-\delta k_{10})V_{11} - \delta^2(1-k_{10})V_{10} \geq 0$, then $\arg\max_{k\in[0,1]} V_{11} = 1$ and $\arg\max_{k\in[0,1]} V_{10} = 1$. By symmetry, $k_{11} = 1$ and $k_{10} = 1$. Hence, the condition can be rewritten

$$
\begin{cases}
(\frac{1}{2}f_\alpha - \delta)v - c^* + \delta(1-\delta)V_{11} \geq 0 \\
(f_\alpha - \delta)v - c_{inf} + \delta(1-\delta)V_{11} \geq 0
\end{cases}
$$

$$
\begin{aligned}
&\Leftrightarrow \delta(1-\delta)V_{11} \geq \max\{c^* - (\frac{1}{2}f_\alpha - \delta)v, c_{inf} - (f_\alpha - \delta)v\} \\
&\Leftrightarrow \delta(1-\delta)V_{11} \geq -(f_\alpha - \delta)v + \max\{c^* + \frac{1}{2}f_\alpha v, c_{inf}\} \\
&\Leftrightarrow \delta(1-\delta)V_{11} \geq -(f_\alpha - \delta)v + c^* + \frac{1}{2}f_\alpha v \\
&\Leftrightarrow \delta(1-\delta)V_{11} \geq c^* - (\frac{1}{2}f_\alpha - \delta)v
\end{aligned}
$$

I have

$$\begin{aligned}
V_{11} =& \frac{1}{2} f_\alpha v - c^* + \delta V_{11} \\
\Leftrightarrow V_{11} =& \frac{1}{1-\delta}\Big(\frac{1}{2} f_\alpha v - c^*\Big)
\end{aligned}$$

Hence, the condition can be rewritten

$$\begin{aligned}
&\delta\Big(\frac{1}{2} f_\alpha v - c^*\Big) \geq c^* - (\frac{1}{2} f_\alpha - \delta)v \\
&\Leftrightarrow ((1+\delta)\frac{1}{2} f_\alpha - \delta)v \geq (1+\delta)c^* \\
&\Leftrightarrow v \geq \frac{c^*}{\frac{1}{2} f_\alpha - \frac{\delta}{1+\delta}}
\end{aligned}$$

In this case, both scientists are never constrained and always choose CS.

If $v \geq \frac{4c_{inf}}{f_\alpha}$, $c^* = \frac{f_\alpha}{4} v$ and the condition becomes

$$\begin{aligned}
v \geq \frac{\frac{f_\alpha}{4} v}{\frac{1}{2} f_\alpha - \frac{\delta}{1+\delta}} \Leftrightarrow & \frac{1}{2} f_\alpha - \frac{\delta}{1+\delta} \geq \frac{f_\alpha}{4} \\
\Leftrightarrow & f_\alpha \geq 4\frac{\delta}{1+\delta}
\end{aligned}$$

If $v < \frac{4c_{inf}}{f_\alpha}$, $c^* = c_{inf}$ and the condition becomes

$$v \geq \frac{c_{inf}}{\frac{1}{2} f_\alpha - \frac{\delta}{1+\delta}}$$

Consequently, the condition for CS with a minimal investment is

$$\frac{c_{inf}}{\frac{1}{2}f_\alpha - \frac{\delta}{1+\delta}} \leq v < \frac{4c_{inf}}{f_\alpha}$$

In this way, a necessary condition for the existence of this case is

$$\frac{c_{inf}}{\frac{1}{2}f_\alpha - \frac{\delta}{1+\delta}} < \frac{4c_{inf}}{f_\alpha}$$
$$\Leftrightarrow 4\frac{\delta}{1+\delta} < f_\alpha$$

Besides, $f_\alpha < 1$. Hence, another necessary condition for this case to exist is

$$4\frac{\delta}{1+\delta} < 1 \Leftrightarrow \delta < \frac{1}{3}$$

2. If $k_{11}\Big((\frac{1}{2}f_\alpha - \delta)v - c^* + \delta(1-\delta k_{10})V_{11} - \delta^2(1-k_{10})V_{10}\Big) + (1-k_{11})\Big((f_\alpha - \delta)v - c_{inf} + \delta V_{10} - \delta^2 V_{11}\Big) \geq 0$ and $(f_\alpha - \delta)v - c_{inf} + \delta(1-\delta k_{10})V_{11} - \delta^2(1-k_{10})V_{10} < 0$, then $\arg\max_{k\in[0,1]} V_{11} = 1$ and $\arg\max_{k\in[0,1]} V_{10} = 0$. By symmetry, $k_{11} = 1$ and $k_{10} = 0$. Hence, the condition can be rewritten

$$\begin{cases} (\frac{1}{2}f_\alpha - \delta)v - c^* + \delta V_{11} - \delta^2 V_{10} \geq 0 \\ (f_\alpha - \delta)v - c_{inf} + \delta V_{11} - \delta^2 V_{10} < 0 \end{cases}$$
$$\Leftrightarrow c^* - (\frac{1}{2}f_\alpha - \delta)v \leq \delta V_{11} - \delta^2 V_{10} < c_{inf} - (f_\alpha - \delta)v$$

For this condition to hold, it must be that

$$
\begin{aligned}
& c^* - (\frac{1}{2}f_\alpha - \delta)v < c_{inf} - (f_\alpha - \delta)v \\
& c^* - c_{inf} < -\frac{1}{2}f_\alpha v
\end{aligned}
$$

which is never the case. Hence, this case does not exist.

3. If $k_{11}\Big((\frac{1}{2}f_\alpha - \delta)v - c^* + \delta(1-\delta k_{10})V_{11} - \delta^2(1-k_{10})V_{10}\Big) + (1-k_{11})\Big((f_\alpha - \delta)v - c_{inf} + \delta V_{10} - \delta^2 V_{11}\Big) < 0$ and $(f_\alpha - \delta)v - c_{inf} + \delta(1-\delta k_{10})V_{11} - \delta^2(1-k_{10})V_{10} \geq 0$, then $\arg\max_{k\in[0,1]} V_{11} = 0$ and $\arg\max_{k\in[0,1]} V_{10} = 1$. By symmetry, $k_{11} = 0$ and $k_{10} = 1$. Hence, the condition can be rewritten

$$
\begin{cases}
(f_\alpha - \delta)v - c_{inf} + \delta V_{10} - \delta^2 V_{11} < 0 \\
(f_\alpha - \delta)v - c_{inf} + \delta(1-\delta)V_{11} \geq 0
\end{cases}
$$

$$
\Leftrightarrow -\delta(1-\delta)V_{11} \leq (f_\alpha - \delta)v - c_{inf} < -\delta(V_{10} - \delta V_{11})
$$

For this condition to hold, it must be that

$$
\begin{aligned}
-\delta(1-\delta)V_{11} < -\delta(V_{10} - \delta V_{11}) \\
\Leftrightarrow V_{11} > V_{10}
\end{aligned}
$$

I have

$$
\begin{aligned}
V_{11} =& \delta(v + \delta V_{11}) \Leftrightarrow V_{11} = \frac{\delta}{1-\delta^2}v \\
V_{10} =& f_\alpha v - c_{inf} + \delta V_{11} = f_\alpha v - c_{inf} + \frac{\delta^2}{1-\delta^2}v
\end{aligned}
$$

The necessary condition can be rewritten

$$\begin{aligned}
&\frac{\delta}{1-\delta^2}v > f_\alpha v - c_{inf} + \frac{\delta^2}{1-\delta^2}v \\
\Leftrightarrow &c_{inf} > \Big(f_\alpha - \frac{\delta(1-\delta)}{1-\delta^2}\Big)v \\
\Leftrightarrow &c_{inf} > \Big(f_\alpha - \frac{\delta}{1+\delta}\Big)v
\end{aligned}$$

Added to the first condition, I must have

$$\begin{aligned}
&\begin{cases}
-\delta(1-\delta)\dfrac{\delta}{1-\delta^2}v \leq (f_\alpha-\delta)v - c_{inf} < -\delta(f_\alpha v - c_{inf} + \dfrac{\delta^2}{1-\delta^2}v - \delta\dfrac{\delta}{1-\delta^2}v) \\
c_{inf} > \Big(f_\alpha - \dfrac{\delta}{1+\delta}\Big)v
\end{cases} \\
\Leftrightarrow &\begin{cases}
-\dfrac{\delta^2}{1+\delta}v \leq (f_\alpha-\delta)v - c_{inf} < -\delta f_\alpha v + \delta c_{inf} \\
c_{inf} > \Big(f_\alpha - \dfrac{\delta}{1+\delta}\Big)v
\end{cases} \\
\Leftrightarrow &\begin{cases}
-\dfrac{\delta^2}{1+\delta}v \leq (f_\alpha-\delta)v - c_{inf} < -\delta f_\alpha v + \delta c_{inf} \\
c_{inf} > \Big(f_\alpha - \dfrac{\delta}{1+\delta}\Big)v
\end{cases} \\
\Leftrightarrow &\begin{cases}
-\dfrac{\delta^2}{1+\delta}v \leq (f_\alpha-\delta)v - c_{inf} \\
(f_\alpha-\delta)v - c_{inf} < -\delta f_\alpha v + \delta c_{inf} \\
c_{inf} > \Big(f_\alpha - \dfrac{\delta}{1+\delta}\Big)v
\end{cases} \\
\Leftrightarrow &\begin{cases}
c_{inf} \leq (f_\alpha - \dfrac{\delta}{1+\delta})v \\
(f_\alpha-\delta)v - c_{inf} < -\delta f_\alpha v + \delta c_{inf} \\
c_{inf} > \Big(f_\alpha - \dfrac{\delta}{1+\delta}\Big)v
\end{cases}
\end{aligned}$$

From the first and second lines, this case does not exist.

4. If $k_{11}\Big((\frac{1}{2}f_\alpha - \delta)v - c^* + \delta(1-\delta k_{10})V_{11} - \delta^2(1-k_{10})V_{10}\Big) + (1-k_{11})\Big((f_\alpha-\delta)v - c_{inf} + \delta V_{10} - \delta^2 V_{11}\Big) < 0$ and $(f_\alpha-\delta)v - c_{inf} + \delta(1-\delta k_{10})V_{11} - \delta^2(1-k_{10})V_{10} < 0$, then

$\arg\max_{k\in[0,1]} V_{11} = 0$ and $\arg\max_{k\in[0,1]} V_{10} = 0$. By symmetry, $k_{11} = 0$ and $k_{10} = 0$. Hence, the condition can be rewritten

$$\begin{cases} (f_\alpha - \delta)v - c_{inf} + \delta V_{10} - \delta^2 V_{11} < 0 \\ (f_\alpha - \delta)v - c_{inf} + \delta V_{11} - \delta^2 V_{10} < 0 \end{cases}$$
$$\Leftrightarrow (f_\alpha - \delta)v - c_{inf} < \delta \min\{\delta V_{11} - V_{10}, \delta V_{10} - V_{11}\}$$

I have

$$\begin{aligned} V_{11} =& \delta\Big(v + \delta V_{11}\Big) && \Leftrightarrow V_{11} = \frac{\delta}{1-\delta^2}v \\ V_{10} =& \delta\Big(v + \delta V_{10}\Big) && \Leftrightarrow V_{10} = \frac{\delta}{1-\delta^2}v \end{aligned}$$

Hence, the condition can be rewritten

$$\begin{aligned} &(f_\alpha - \delta)v - c_{inf} < \delta(\delta - 1)\frac{\delta}{1-\delta^2}v \\ \Leftrightarrow&(f_\alpha - \delta)v - c_{inf} < -\frac{\delta^2}{1+\delta}v \\ \Leftrightarrow&(f_\alpha - \frac{\delta}{1+\delta})v < c_{inf} \end{aligned}$$

This condition is the same as the monopolist case. Hence, both scientists choose TS if $v < v^*(f_\alpha)$.

Proof of Proposition 4

There are 5 cases.

1. If $k_{11}\Big((\frac{1}{2}f_\alpha - \delta)v - c^* + \delta(1-\delta k_{10})V_{11} - \delta^2(1-k_{10})V_{10}\Big) + (1-k_{11})\Big((f_\alpha - \delta)v - c_{inf} + \delta V_{10} - \delta^2 V_{11}\Big) \geq 0$ and $(f_\alpha - \delta)v - c_{inf} + \delta(1-\delta k_{10})V_{11} - \delta^2(1-k_{10})V_{10} = 0$, then $\arg\max_{k\in[0,1]} V_{11} = 1$ and $\arg\max_{k\in[0,1]} V_{10} = k \in (0,1)$. By symmetry, $k_{11} = 1$ and

$k_{10} \in (0,1)$. Hence, the condition can be rewritten

$$\begin{cases} (\frac{1}{2}f_\alpha - \delta)v - c^* + \delta(1-\delta k_{10})V_{11} - \delta^2(1-k_{10})V_{10} \geq 0 \\ (f_\alpha - \delta)v - c_{inf} + \delta(1-\delta k_{10})V_{11} - \delta^2(1-k_{10})V_{10} = 0 \end{cases}$$
$$\Leftrightarrow (\frac{1}{2}f_\alpha - \delta)v - c^* \geq (f_\alpha - \delta)v - c_{inf}$$
$$\Leftrightarrow c_{inf} - c^* \geq \frac{1}{2}v$$
$$\Leftrightarrow c_{inf} - \max\{\frac{f_\alpha v}{4}, c_{inf}\} \geq \frac{1}{2}v$$

which is never true. Hence, this case never occurs.

2. If $k_{11}\Big((\frac{1}{2}f_\alpha - \delta)v - c^* + \delta(1-\delta k_{10})V_{11} - \delta^2(1-k_{10})V_{10}\Big) + (1-k_{11})\Big((f_\alpha - \delta)v - c_{inf} + \delta V_{10} - \delta^2 V_{11}\Big) < 0$ and $(f_\alpha - \delta)v - c_{inf} + \delta(1-\delta k_{10})V_{11} - \delta^2(1-k_{10})V_{10} = 0$, then $\arg\max_{k\in[0,1]} V_{11} = 0$ and $\arg\max_{k\in[0,1]} V_{10} = k \in (0,1)$. By symmetry, $k_{11} = 0$ and $k_{10} \in (0,1)$. Hence, the conditions can be rewritten

$$\begin{cases} (f_\alpha - \delta)v - c_{inf} + \delta V_{10} - \delta^2 V_{11} < 0 \\ (f_\alpha - \delta)v - c_{inf} + \delta(1-\delta k_{10})V_{11} - \delta^2(1-k_{10})V_{10} = 0 \end{cases}$$
$$\Leftrightarrow \begin{cases} (f_\alpha - \delta)v - c_{inf} + \delta V_{10} - \delta^2 V_{11} < 0 \\ (f_\alpha - \delta)v - c_{inf} + \delta V_{11} - \delta^2 V_{10} + \delta^2 k_{10}(V_{10} - V_{11}) = 0 \end{cases}$$

From the equality, I have

$$k_{10} = \frac{(f_\alpha - \delta)v - c_{inf} + \delta V_{11} - \delta^2 V_{10}}{\delta^2(V_{11} - V_{10})}$$

Besides, $k_{10} \in (0,1)$, hence

$$\begin{cases} \dfrac{(f_\alpha - \delta)v - c_{inf} + \delta V_{11} - \delta^2 V_{10}}{\delta^2(V_{11} - V_{10})} > 0 \\ (f_\alpha - \delta)v - c_{inf} + \delta V_{11} - \delta^2 V_{10} < \delta^2(V_{11} - V_{10}) \end{cases}$$

Moreover, replacing $k_{11} = 0$ and k_{10} in V_{11} and V_{10}, I obtain

$$\begin{aligned} V_{11} =& \delta(v + \delta V_{11}) \Leftrightarrow V_{11} = \frac{\delta}{1-\delta^2} v \\ V_{10} =& k_{10}\Big[f_\alpha v - c_{inf} + \delta V_{11}\Big] + (1-k_{10})\delta\Big[v + \delta\Big(k_{10}V_{11} + (1-k_{10})V_{10}\Big)\Big] \\ \Leftrightarrow V_{10} =& f_\alpha v - c_{inf} + \frac{\delta^2}{1-\delta^2} v \end{aligned}$$

So this case exists if 3 conditions are reunited:

$$\begin{cases} (f_\alpha - \delta)v - c_{inf} + \delta V_{10} - \delta^2 V_{11} < 0 \\ \dfrac{(f_\alpha - \delta)v - c_{inf} + \delta V_{11} - \delta^2 V_{10}}{\delta^2(V_{11} - V_{10})} > 0 \\ (f_\alpha - \delta)v - c_{inf} + \delta V_{11} - \delta^2 V_{10} < \delta^2(V_{11} - V_{10}) \end{cases}$$

$$\Leftrightarrow \begin{cases} \Big(f_\alpha - \dfrac{\delta}{1+\delta}\Big)v - c_{inf} < 0 \\ \dfrac{(1-\delta^2)(f_\alpha v - c_{inf}) - \delta(1-\delta)v}{\delta^2(\frac{\delta}{1+\delta}v - f_\alpha v + c_{inf})} > 0 \\ \Big(f_\alpha - \dfrac{\delta}{1+\delta}\Big)v - c_{inf} < 0 \end{cases}$$

$$\Leftrightarrow \begin{cases} \Big(f_\alpha - \dfrac{\delta}{1+\delta}\Big)v - c_{inf} < 0 \\ \Big(f_\alpha - \dfrac{\delta}{1+\delta}\Big)v - c_{inf} > 0 \end{cases}$$

These 2 inequalities cannot be true at the same time, hence this case never occurs.

3. If $k_{11}\Big((\frac{1}{2}f_\alpha - \delta)v - c^* + \delta(1-\delta k_{10})V_{11} - \delta^2(1-k_{10})V_{10}\Big) + (1-k_{11})\Big((f_\alpha - \delta)v - c_{inf} + \delta V_{10} - \delta^2 V_{11}\Big) = 0$ and $(f_\alpha - \delta)v - c_{inf} + \delta(1-\delta k_{10})V_{11} - \delta^2(1-k_{10})V_{10} \geq 0$, then

$\arg\max_{k\in[0,1]} V_{11} = k \in (0,1)$ and $\arg\max_{k\in[0,1]} V_{10} = 1$. By symmetry, $k_{11} \in (0,1)$ and $k_{10} = 1$. Hence, the condition can be rewritten

$$\begin{cases} k_{11}\Big((\frac{1}{2}f_\alpha - \delta)v - c^* + \delta(1-\delta)V_{11}\Big) + (1-k_{11})\Big((f_\alpha - \delta)v - c_{inf} + \delta V_{10} - \delta^2 V_{11}\Big) = 0 \\ (f_\alpha - \delta)v - c_{inf} + \delta(1-\delta)V_{11} \geq 0 \end{cases}$$
$$\Leftrightarrow \begin{cases} (f_\alpha - \delta)v - c_{inf} - \delta^2 V_{11} + \delta V_{10} + k_{11}\Big(c_{inf} - c^* - \frac{1}{2}f_\alpha v + \delta V_{11} - \delta V_{10}\Big) = 0 \\ (f_\alpha - \delta)v - c_{inf} + \delta(1-\delta)V_{11} \geq 0 \end{cases}$$

Hence

$$k_{11} = \frac{(f_\alpha - \delta)v - c_{inf} + \delta(V_{10} - \delta V_{11})}{c^* - c_{inf} + \frac{1}{2}f_\alpha v + \delta(V_{10} - V_{11})}$$

Since k_{11} must lie in $(0,1)$, I have 3 inequalities

$$\begin{cases} (f_\alpha - \delta)v - c_{inf} + \delta(V_{10} - \delta V_{11}) < c^* - c_{inf} + \frac{1}{2}f_\alpha v + \delta(V_{10} - V_{11}) \\ \dfrac{(f_\alpha - \delta)v - c_{inf} + \delta(V_{10} - \delta V_{11})}{c^* - c_{inf} + \frac{1}{2}f_\alpha v + \delta(V_{10} - V_{11})} > 0 \\ (f_\alpha - \delta)v - c_{inf} + \delta(1-\delta)V_{11} \geq 0 \end{cases}$$
$$\Leftrightarrow \begin{cases} c_{inf} \leq (f_\alpha - \delta)v + \delta(1-\delta)V_{11} < c^* + \frac{1}{2}f_\alpha v \\ \dfrac{(f_\alpha - \delta)v - c_{inf} + \delta(V_{10} - \delta V_{11})}{c^* - c_{inf} + \frac{1}{2}f_\alpha v + \delta(V_{10} - V_{11})} > 0 \end{cases}$$

I have

$$\begin{aligned} V_{11} =& \delta(v + \delta V_{11}) \Leftrightarrow V_{11} = \frac{\delta}{1-\delta^2}v \\ V_{10} =& f_\alpha v - c_{inf} + \delta V_{11} \\ =& f_\alpha v - c_{inf} + \frac{\delta^2}{1-\delta^2}v \end{aligned}$$

So the system of inequalities becomes

$$\begin{cases} c_{inf} \le (f_\alpha - \delta)v + \dfrac{\delta^2}{1+\delta}v < c^* + \dfrac{1}{2}f_\alpha v \\ \dfrac{(f_\alpha - \delta)v - c_{inf} + \delta(f_\alpha v - c_{inf})}{c^* - c_{inf} + \dfrac{1}{2}f_\alpha v + \delta(f_\alpha v - c_{inf} - \dfrac{\delta}{1+\delta}v)} > 0 \end{cases}$$

$$\Leftrightarrow \begin{cases} c_{inf} \le (f_\alpha - \dfrac{\delta}{1+\delta})v < c^* + \dfrac{1}{2}f_\alpha v \\ (1+\delta)\dfrac{(f_\alpha - \dfrac{\delta}{1+\delta})v - c_{inf}}{c^* - c_{inf} + \dfrac{1}{2}f_\alpha v + \delta(f_\alpha v - c_{inf} - \dfrac{\delta}{1+\delta}v)} > 0 \end{cases}$$

$$\Leftrightarrow \begin{cases} (f_\alpha - \dfrac{\delta}{1+\delta})v \ge c_{inf} \\ (\dfrac{f_\alpha}{2} - \dfrac{\delta}{1+\delta})v < c^* \\ \Big[(\dfrac{1}{2}+\delta)f_\alpha - \dfrac{\delta^2}{1+\delta}\Big]v > (1+\delta)c_{inf} - c^* \end{cases}$$

$$\Leftrightarrow \begin{cases} (f_\alpha - \dfrac{\delta}{1+\delta})v \ge c_{inf} \\ \dfrac{f_\alpha}{4}v \ge c_{inf} \\ (\dfrac{f_\alpha}{4} - \dfrac{\delta}{1+\delta})v < 0 \\ \Big[(\dfrac{3}{4}+\delta)f_\alpha - \dfrac{\delta^2}{1+\delta}\Big]v > (1+\delta)c_{inf} \end{cases} \text{or} \begin{cases} (f_\alpha - \dfrac{\delta}{1+\delta})v \ge c_{inf} \\ \dfrac{f_\alpha}{4}v < c_{inf} \\ (\dfrac{f_\alpha}{2} - \dfrac{\delta}{1+\delta})v < c_{inf} \\ \Big[(\dfrac{1}{2}+\delta)f_\alpha - \dfrac{\delta^2}{1+\delta}\Big]v > \delta c_{inf} \end{cases}$$

$$\Leftrightarrow \begin{cases} v \ge \dfrac{c_{inf}}{f_\alpha - \dfrac{\delta}{1+\delta}} \\ v \ge \dfrac{4c_{inf}}{f_\alpha} \\ \dfrac{\delta}{1+\delta} < f_\alpha < \dfrac{4\delta}{1+\delta} \\ v > \dfrac{(1+\delta)c_{inf}}{(\dfrac{3}{4}+\delta)f_\alpha - \dfrac{\delta^2}{1+\delta}} \end{cases} \text{or} \begin{cases} v \ge \dfrac{c_{inf}}{f_\alpha - \dfrac{\delta}{1+\delta}} \\ v < \dfrac{4c_{inf}}{f_\alpha} \\ f_\alpha > \dfrac{\delta}{1+\delta} \\ (\dfrac{f_\alpha}{2} - \dfrac{\delta}{1+\delta})v < c_{inf} \end{cases}$$

$$\Leftrightarrow \begin{cases} v \ge \dfrac{c_{inf}}{f_\alpha - \dfrac{\delta}{1+\delta}} \\ v \ge \dfrac{4c_{inf}}{f_\alpha} \\ \dfrac{\delta}{1+\delta} < f_\alpha < \dfrac{4\delta}{1+\delta} \end{cases} \text{or} \begin{cases} v \ge \dfrac{c_{inf}}{f_\alpha - \dfrac{\delta}{1+\delta}} \\ v < \dfrac{4c_{inf}}{f_\alpha} \\ f_\alpha > \dfrac{\delta}{1+\delta} \\ (\dfrac{f_\alpha}{2} - \dfrac{\delta}{1+\delta})v < c_{inf} \end{cases}$$

$$\Leftrightarrow \begin{cases} v \ge \max\{v^*(f_\alpha), \tilde{v}_1(f_\alpha)\} \\ f_\alpha < \tilde{f}_\alpha \end{cases} \text{or } v^*(f_\alpha) \le v < \min\{\tilde{v}_1(f_\alpha), \tilde{v}_2(f_\alpha)\}$$

4. If $k_{11}\Big((\frac{1}{2}f_\alpha - \delta)v - c^* + \delta(1-\delta k_{10})V_{11} - \delta^2(1-k_{10})V_{10}\Big) + (1-k_{11})\Big((f_\alpha - \delta)v - c_{inf} + \delta V_{10} - \delta^2 V_{11}\Big) = 0$ and $(f_\alpha - \delta)v - c_{inf} + \delta(1-\delta k_{10})V_{11} - \delta^2(1-k_{10})V_{10} < 0$, then $\arg\max_{k\in[0,1]} V_{11} = k \in (0,1)$ and $\arg\max_{k\in[0,1]} V_{10} = 0$. By symmetry, $k_{11} \in (0,1)$ and $k_{10} = 0$. Hence, the condition can be rewritten

$$\begin{cases} k_{11}\Big((\frac{1}{2}f_\alpha - \delta)v - c^* + \delta V_{11} - \delta^2 V_{10}\Big) + (1-k_{11})\Big((f_\alpha - \delta)v - c_{inf} + \delta V_{10} - \delta^2 V_{11}\Big) = 0 \\ (f_\alpha - \delta)v - c_{inf} + \delta V_{11} - \delta^2 V_{10} < 0 \end{cases}$$
$$\Leftrightarrow \begin{cases} (f_\alpha - \delta)v - c_{inf} + \delta V_{10} - \delta^2 V_{11} + k_{11}\Big(c_{inf} - c^* - \frac{1}{2}f_\alpha v + \delta(1+\delta)(V_{11} - V_{10})\Big) = 0 \\ (f_\alpha - \delta)v - c_{inf} + \delta V_{11} - \delta^2 V_{10} < 0 \end{cases}$$

From the equality, I have

$$k_{11} = \frac{(f_\alpha - \delta)v - c_{inf} + \delta V_{10} - \delta^2 V_{11}}{c^* - c_{inf} + \frac{1}{2}f_\alpha v + \delta(1+\delta)(V_{10} - V_{11})} \in (0,1)$$

I have

$$\begin{aligned} V_{10} =& \delta\Big(v + \delta V_{10}\Big) \Leftrightarrow V_{10} = \frac{\delta}{1-\delta^2}v \\ V_{11} =& k_{11}\delta\Big(v + \delta V_{10}\Big) + (1-k_{11})\delta(v + \delta V_{11}) \\ =& k_{11}\frac{\delta}{1-\delta^2}v + (1-k_{11})\delta(v + \delta V_{11}) \end{aligned}$$

There are 2 solutions for V_{11}:

$$\begin{aligned} V_{11}^1 =& \frac{\delta}{1-\delta^2}v \\ V_{11}^2 =& \frac{f_\alpha(1-\delta^4)v + 2\delta^2 v + 2(1-\delta^2)\Big((1-\delta^2)c^* - c_{inf}\Big)}{2\delta(1+\delta-2\delta^2-\delta^3+\delta^4)} \end{aligned}$$

If $V_{11} = V_{11}^1$, then

$$k_{11} = \frac{(f_\alpha - \frac{\delta}{1+\delta})v - c_{inf}}{c^* - c_{inf} + \frac{1}{2} f_\alpha v}$$

Since $k_{11} \in (0,1)$, I have

$$\begin{cases} (f_\alpha - \frac{\delta}{1+\delta})v > c_{inf} \\ (f_\alpha - \frac{\delta}{1+\delta})v < c_{inf} \end{cases}$$

which is impossible, hence V_{11} cannot be equal to V_{11}^1. If $V_{11} = V_{11}^2$, then

$$k_{11} = -\frac{1-\delta^2}{\delta} < 0$$

Hence, $k_{11} \notin (0,1)$ and this case does not exist.

5. If $k_{11}\Big((\frac{1}{2}f_\alpha - \delta)v - c^* + \delta(1-\delta k_{10})V_{11} - \delta^2(1-k_{10})V_{10}\Big) + (1-k_{11})\Big((f_\alpha - \delta)v - c_{inf} + \delta V_{10} - \delta^2 V_{11}\Big) = 0$ and $(f_\alpha - \delta)v - c_{inf} + \delta(1-\delta k_{10})V_{11} - \delta^2(1-k_{10})V_{10} = 0$, then $\arg\max_{k\in[0,1]} V_{11} = k \in [0,1]$ and $\arg\max_{k\in[0,1]} V_{10} = k \in (0,1)$. By symmetry,

$k_{11} \in (0,1)$ and $k_{10} \in [0,1]$. Hence, the condition can be rewritten

$$\begin{cases} k_{11}\big((\frac{1}{2}f_\alpha - \delta)v - c^* + \delta(1-\delta k_{10})V_{11} - \delta^2(1-k_{10})V_{10}\big) + (1-k_{11})\big((f_\alpha - \delta)v - c_{inf} + \delta V_{10} - \delta^2 V_{11}\big) = 0 \\ (f_\alpha - \delta)v - c_{inf} + \delta(1-\delta k_{10})V_{11} - \delta^2(1-k_{10})V_{10} = 0 \end{cases}$$

$$\Leftrightarrow \begin{cases} k_{11}\left(c_{inf} - c^* - \frac{1}{2}f_\alpha v + \delta(1+\delta(1-k_{10}))(V_{11}-V_{10})\right) + (f_\alpha - \delta)v - c_{inf} + \delta(V_{10} - \delta V_{11}) = 0 \\ k_{10} = \dfrac{(f_\alpha - \delta)v - c_{inf} + \delta(V_{11} - \delta V_{10})}{\delta^2(V_{11} - V_{10})} \end{cases}$$

$$\Leftrightarrow \begin{cases} k_{11} = \dfrac{(f_\alpha - \delta)v - c_{inf} + \delta(V_{10} - \delta V_{11})}{c^* - c_{inf} + \frac{1}{2}f_\alpha v + \delta(1+\delta(1-k_{10}))(V_{10}-V_{11})} \\ k_{10} = \dfrac{(f_\alpha - \delta)v - c_{inf} + \delta(V_{11} - \delta V_{10})}{\delta^2(V_{11} - V_{10})} \end{cases}$$

$$\Rightarrow k_{11} = \frac{(f_\alpha - \delta)v - c_{inf} + \delta(V_{10} - \delta V_{11})}{c^* - c_{inf} + \frac{1}{2}f_\alpha v + \delta\left(1 + \dfrac{c_{inf} - (f_\alpha - \delta)v - \delta(1-\delta)V_{11}}{\delta(V_{11} - V_{10})}\right)(V_{10} - V_{11})}$$

$$= \frac{(f_\alpha - \delta)v - c_{inf} + \delta(V_{10} - \delta V_{11})}{c^* - c_{inf} + \frac{1}{2}f_\alpha v - c_{inf} + (f_\alpha - \delta)v + \delta(V_{10} - \delta V_{11})}$$

k_{11} and k_{10} must lie in $(0,1)$, hence

$$\begin{cases} c^* - c_{inf} + \frac{1}{2}f_\alpha v - c_{inf} + (f_\alpha - \delta)v + \delta(V_{10} - \delta V_{11}) > c^* - c_{inf} + \frac{1}{2}f_\alpha v \\ (f_\alpha - \delta)v - c_{inf} + \delta(V_{10} - \delta V_{11}) < c^* - c_{inf} + \frac{1}{2}f_\alpha v - c_{inf} + (f_\alpha - \delta)v + \delta(V_{10} - \delta V_{11}) \\ \dfrac{(f_\alpha - \delta)v - c_{inf} + \delta(V_{11} - \delta V_{10})}{\delta^2(V_{11} - V_{10})} > 0 \\ (f_\alpha - \delta)v - c_{inf} + \delta(V_{11} - \delta V_{10}) < \delta^2(V_{11} - V_{10}) \end{cases}$$

$$\Leftrightarrow \begin{cases} (f_\alpha - \delta)v - c_{inf} + \delta(V_{10} - \delta V_{11}) > 0 \\ c^* - c_{inf} + \frac{1}{2}f_\alpha v > 0 \\ \dfrac{(f_\alpha - \delta)v - c_{inf} + \delta(V_{11} - \delta V_{10})}{\delta^2(V_{11} - V_{10})} > 0 \\ (f_\alpha - \delta)v - c_{inf} + \delta(1 - \delta)V_{11} < 0 \end{cases}$$

$$\Leftrightarrow \begin{cases} -\delta(1 - \delta)V_{11} > (f_\alpha - \delta)v - c_{inf} > \delta(\delta V_{11} - V_{10}) \\ V_{11} < V_{10} \\ (f_\alpha - \delta)v - c_{inf} + \delta(V_{11} - \delta V_{10}) < 0 \end{cases}$$

$$\Leftrightarrow \begin{cases} \delta \max\{V_{11} - \delta V_{10}, (1 - \delta)V_{11}\} < c_{inf} - (f_\alpha - \delta)v < \delta(V_{10} - \delta V_{11}) \\ V_{11} < V_{10} \end{cases}$$

$$\Leftrightarrow \begin{cases} \delta(1 - \delta)V_{11} < c_{inf} - (f_\alpha - \delta)v < \delta(V_{10} - \delta V_{11}) \\ V_{11} < V_{10} \end{cases}$$

I have

$$
\begin{aligned}
V_{10} =& \delta\big[v + \delta\big(k_{10}V_{11} + (1-k_{10})V_{10}\big)\big] \\
=& \delta\bigg[v + \frac{1}{\delta(V_{11}-V_{10})}\Big(((f_\alpha-\delta)v - c_{inf} + \delta(V_{11}-\delta V_{10}))V_{11} + \big(c_{inf} - (f_\alpha-\delta)v - \delta(1-\delta)V_{11}\big)V_{10}\Big)\bigg] \\
=& \frac{1}{V_{11}-V_{10}}\bigg[\delta(V_{11}-V_{10})v + ((f_\alpha-\delta)v - c_{inf})(V_{11}-V_{10}) + \delta(V_{11}-\delta V_{10})V_{11} - \delta(1-\delta)V_{11}V_{10}\Big)\bigg] \\
=& f_\alpha v - c_{inf} + \delta V_{11} \\
V_{11} =& k_{11}\delta\big[v + \delta\big(k_{10}V_{11} + (1-k_{10})V_{10}\big)\big] + (1-k_{11})\delta(v + \delta V_{11}) \\
=& k_{11}\big(f_\alpha v - c_{inf} + \delta V_{11}\big) + (1-k_{11})\delta(v + \delta V_{11})
\end{aligned}
$$

Then

$$
\begin{aligned}
& V_{10} - V_{11} = (1-k_{11})\Big((f_\alpha-\delta)v - c_{inf} + \delta(1-\delta)V_{11}\Big) > 0 \\
& \Rightarrow \delta(1-\delta)V_{11} > c_{inf} - (f_\alpha-\delta)v
\end{aligned}
$$

which is incompatible with the conditions, hence this case never exists.

Bibliography

Bagnoli, Mark and Barton L. Lipman (1989) "Provision of public goods: Fully implementing the core through private contributions," *The Review of Economic Studies*, 56 (4), 583–601.

Bergstrom, Theodore, Lawrence Blume, and Hal Varian (1986) "On the private provision of public goods," *Journal of public economics*, 29 (1), 25–49.

Bernard, Olga (2020) "Fighting for priority: researcher's dilemma with an alternative," working paper.

Bramoullé, Yann and Gilles Saint-Paul (2010) "Research cycles," *Journal of economic theory*, 145 (5), 1890–1920.

Cooper, Caren B., Jennifer Shirk, and Benjamin Zuckerberg (2014) "The invisible prevalence of citizen science in global research: migratory birds and climate change," *PloS one*, 9 (9), e106508.

Franzoni, Chiara and Henry Sauermann (2014) "Crowd science: The organization of scientific research in open collaborative projects," *Research policy*, 43 (1), 1–20.

Garfagnini, Umberto and Bruno Strulovici (2012) "Social learning and innovation cycles,"Technical report, CSIO Working Paper.

Koeze, Ella and Nathaniel Popper (2020) "The virus changed the way we internet," *The New York Times*.

Kullenberg, Christopher and Dick Kasperowski (2016) "What is citizen science?–A scientometric meta-analysis," *PloS one*, 11 (1), e0147152.

Laut, Jeffrey, Francesco Cappa, Oded Nov, and Maurizio Porfiri (2017) "Increasing citizen science contribution using a virtual peer," *Journal of the Association for Information Science and Technology*, 68 (3), 583–593.

Lukyanenko, Roman, Jeffrey Parsons, and Yolanda F. Wiersma (2016) "Emerging problems of data quality in citizen science," *Conservation Biology*, 30 (3), 447–449.

Mahajan, Sachit, Prashant Kumar, Janaina Antonino Pinto, Agnese Riccetti, Katinka Schaaf, Guillem Camprodon, Viktor Smari, Antonella Passani, and Giuseppe Forino (2020) "A citizen science approach for enhancing public understanding of air pollution," *Sustainable Cities and Society*, 52, 101800.

Nov, Oded, Ofer Arazy, and David Anderson (2014) "Scientists@ Home: what drives the quantity and quality of online citizen science participation?" *PloS one*, 9 (4), e90375.

Poisson, Autumn C., Ian M. McCullough, Kendra S. Cheruvelil, Kevin C. Elliott, Jo A. Latimore, and Patricia A. Soranno (2020) "Quantifying the contribution of citizen science to broad-scale ecological databases," *Frontiers in Ecology and the Environment*, 18 (1), 19–26.

Prestopnik, Nathan, Kevin Crowston, and Jun Wang (2017) "Gamers, citizen scientists, and data: Exploring participant contributions in two games with a purpose," *Computers in Human Behavior*, 68, 254–268.

Rowland, Katherine (2012) "Citizen science goes 'extreme'," *Nature News*, 17.

Stokey, Nancy L, RE Lucas, and EC Prescott (1989) "Recursive methods in dynamic economics," *Cambridge, MA: Harvard University*.

Theobald, Ellinore J., Ailene K. Ettinger, Hillary K. Burgess, Lauren B. DeBey, Natalie R. Schmidt, Halley E. Froehlich, Christian Wagner, Janneke Hille Ris Lambers, Joshua

Tewksbury, Melanie A. Harsch et al. (2015) "Global change and local solutions: Tapping the unrealized potential of citizen science for biodiversity research," *Biological Conservation*, 181, 236–244.

Varian, Hal (2004) "System reliability and free riding," in *Economics of information security*, 1–15: Springer.

Watson, David and Luciano Floridi (2018) "Crowdsourced science: sociotechnical epistemology in the e-research paradigm," *Synthese*, 195 (2), 741–764.

Chapter 3

Citizen science and preemption

Abstract I study the tradeoff of the scientist who chooses between citizen science (CS) and traditional science (TS) for her research. After time 0, there exists a new idea which can be studied. Two scientists can discover this idea at any time. Their objective is to publish it before the other one does. However, competition is only potential as a scientist's discovery time is not observable by her competitor. At the scientist's discovery time, she faces two technological choices, which are not available past that time. The first one is TS: she takes time to let the idea mature. There exists an optimal maturation delay which maximizes her publication payoff with TS without any competitor. With the second choice, called CS, the scientist involves citizens' help to publish instantaneously the idea. However, she incurs a fixed cost to make the idea available to citizens. Moreover, the latters are non-experts so there is some risk error that the publication is of bad quality, which brings her no payoff. Focusing on Bayesian pure-strategy equilibria, I prove that there exists two kind of symmetric stationary equilibria. When CS is low-cost, every scientist chooses CS. Otherwise, everyone chooses TS. Besides, I study equilibria in which every scientist chooses TS before a discovery time threshold and CS after. There exists no such equilibrium when it satisfies one of

these two assumptions: i) the threshold discovery time is lower than the equilibrium publication time of the scientist discovering the idea at time 0 and ii) the equilibrium scientists' strategy is continuous at the threshold discovery time. At last, there exists no asymmetric equilibrium in which scientists choose different technological choices.

Keywords: .

JEL codes: .

1 Introduction

In our highly interconnected world, scientists face many challenges to publish their work. To begin with, new methods and techniques are available to work on a novel idea. More specifically, citizen science is a new way of doing science which has emerged this last decade. It enables the scientist to use the citizens' help to work faster than with traditional science. For instance, Galaxy Zoo is a project created in 2007 in which citizens treat over a million of galaxies' data in order to report dust particules. Without the citizens' help, a type of astronomic object would not have been identified and the data treatment would have taken several decades. Another famous citizen science project is FoldIt. Launched in 2008, this online game allows the citizens to work on proteins shapes to better understand their functioning. In 2010, more than 200000 people already participated to that project. In short, citizen science is caracterized by its speed of research compared to traditional science.

Besides, citizen science has no formal structure. In 2009, Timothy Gowers, a 1998 Fields Medal winner mathematician, started to post on his blog a mathematical problem as an experiment to collectively solve it. Within six weeks, several contributors, including university professors, school teachers, PhD students and another Fields medalist, discussed until a proof was found. Eventually, the article stating the proof was published in the Annals of Mathematics under the pseudonym "D.H.J. Polymath" (Polymath, 2012). Thrilled by this success, Timothy Gowers posted several other problems. The structure of this citizen science project differs from FoldIt and Galaxy Zoo as it is rather addressed to mathematicians. Nevertheless, it shares the feature that anybody is able to join the project. Eventually, citizen science enables to implement projects very fast but citizens are not experts. In this respect, the quality of the project outcome is questionable, as stated by Lukyanenko et al. (2016).

Another challenge is inherent to the fact that scientists are aware almost instantaneously of publication around the world. Hence, when a scientist works on a new idea, another one could be working on it at the other side of the planet but she is not aware of it. She only notices this fact when a potential competitor publishes first the idea. As being a "follower" brings only little value, the scientist aims at having the "first-mover" advantage. In this respect, her fear of being preempted in this fight for prioriy leads her to publish her idea before it is mature enough.

To study these challenges, I build a theoretical model using a preemption game approach. At the beginning of the game, a new idea can be discovered by two scientists. Scientists discover the idea at random times, which are independently drawn from an exponential distribution. The discovery time of one scientist is not observed by her competitor. After that time, they can publish the discovery. However, they fight for priority. It means that when one scientist publishes her discovery, her competitor cannot publish and the game stops.

At her discovery time, each scientist faces a technological choice: traditional science (TS) or citizen science (CS). This choice is not available anymore past that time. If she chooses TS, she takes time to let the idea mature. Her publication payoff depends on the idea's maturation delay. Besides, there is an optimal maturation delay which maximizes her payoff without a competitor. If she is preempted by her rival, she enjoys no payoff. I assume that the two scientists cannot publish simultaneously.

If she chooses CS, the scientist involves the help of citizens for her research, which enables her to publish instantaneously her discovery. This reflects the fact that citizen science is faster than traditional science. To make the idea available to citizens, the scientist incurs a fixed cost. As CS involves non-experts, there exists a risk that the publication is of bad quality, which brings her no payoff. I assume that the CS cost is low, so that the scientist's payoff with CS is strictly positive. In this way, there always exists a trade-off between the two technological choices.

Focusing on Bayesian pure-strategy equilibria, a scientist's strategy is composed by her technological choice and her publication time if she chooses TS. I first look at a scientist's strategy if she chooses TS. Her payoff is equal to the publication payoff by letting the idea mature until her publication time, multiplied by the probability that her competitor (he) does not publish before that time, knowing that he has not published before her discovery time. In any Bayesian equilibrium, a scientist who uses TS to publish at her equilibrium publication time must have at the same time a higher payoff than if she chooses TS and publishes after any other time, and than the CS payoff. A scientist who uses CS must have a higher payoff than if she uses TS to publish at any time. First, I focus on symmetric equilibria, in which both scientists have the same strategy. Using a standard revealed-preference argument, I prove that is both scientists choose TS, the equilibrium publication time is weakly increasing in the discovery time. In addition and to facilitate the analysis, I assume that it is strictly increasing.

By first studying symmetric stationary equilibrium, a possible candidate for equilibrium would be that early-born scientists waits before publication a constant maturation delay and late-born choose CS. There would be a threshold discovery time such that a scientist discovering this idea at that time is indifferent between the two technological choices. To study this type of equilibrium, the equilibrium payoff of an early-born scientist who chooses TS must be compared to the CS payoff. Solving the maximization problem of a scientist choosing TS, I find that the equilibrium maturation delay depends on the discovery speed but not on the discovery time. The payoff of a scientist publishing with TS after this stationary maturation delay depends on whether her discovery time is higher or lower than this maturation delay. If it is higher, then the probability of being preempted is the probability that her competitor is born during the interval of a length equal to the maturation delay, before her discovery time. If it is lower, the probability of being preempted is the probability that her competitor discovers the idea before her.

When the CS cost is low, even the scientist discovering the idea at time 0 prefers CS. Hence, at equilibrium, both scientists choose CS. When the CS cost is high, both scientists choose TS at equilibrium. A discontinuous stationary equilibrium may exist when the CS cost is intermediate. I focus on such equilibria, in which the two scientists always publish after the threshold discovery time. As a matter of fact, the scientist discovering the idea at that threshold strictly prefers CS to TS. This is a contradiction to the fact that she must be indifferent between the two technological choices. Hence, such an equilibrium does not exist. Consequently, when the CS cost is intermediate, the two scientists choose CS. In this way, there exist only two kinds of symmetric stationary equilibria. When the CS cost is low, scientists choose CS. Otherwise, they choose TS.

To go further, a discontinuous equilibrium may exist in the intermediate CS cost range if the equilibrium scientists' strategy is non-stationary. Again, I suppose that the two scientists always publish after the threshold discovery time. Scientists discovering the idea before the latter face a low preemption risk, whereas those discovering the idea after the threshold face a high preemtion risk. Combining the incitation constraints of two scientists, one discovering the idea just before the threshold and one just after, I find that the difference between the two payoffs tends to 0 as their discovery times tend to the threshold. Consequently, the equilibrium payoff function is continuous. In this respect, the scientist discovering the idea at the threshold is indifferent between the two technological choices. Similar to the non-existence proof of a stationary equilibrium, a scientist who discovers the idea just before the threshold faces already a too high preemption risk and chooses CS instead. Hence, such a discontinuous non-stationary equilibrium does not exist.

The non-existence problem of an equilibrium in which early-born scientists choose TS and late-borns choose CS may come from the discontinuity at the threshold discovery time. In this respect, I look for a symmetric equilibrium which exhibits a continuous strategy. As the discovery time of an early-born scientist tends to the threshold, her maturation delay tends to 0. However, as her maturation delay goes to 0, so does her payoff with TS. When her discovery time is sufficiently close to the threshold, the payoff with TS is lower than the CS payoff. It implies that she has a strictly profitable deviation by choosing CS. Consequently, there exists no such equilibrium.

Finally, the only type of equilibrium which exists is when all scientists choose the same technological choice, no matter their discovery times. This creates a congestion in the CS use. Indeed, citizens have limited time and attention to get involved in the scientists' discoveries. An equilibrium in which early-born scientists choose TS and late-born ones choose CS would have been interesting as they have time to adapt to the technological change CS. An asymmetric equilibrium would remove the congestion problem with CS, that is, the two scientists make different technological choices. I focus on the case where the scientist who chooses TS publishes after a constant maturation delay. As a matter of fact, an early-born scientist who chooses TS faces a preemption risk, which does not exist in a symmetric equilibrium. Hence, she has a strictly profitable deviation, which is to choose CS as her competitor. Consequently, there exists no such asymmetric equilibrium.

Related literature Close to this chapter lies Bernard (2020). In this paper, I study the researchers' tradeoff between traditional science and citizen science. For any successive project available of a list of projects, two researchers choose between traditional science and citizen science to implement it. On one hand, traditional science takes time but the implementation is certain. On the other hand, citizen science enables her to implement instantaneously the project but with a failure implementation risk. In this respect, my model is in line with Bernard (2020). This risk depends on the effort exerted by two citizens. Besides, scientists fight to attract citizens on their project. By contrast, in this chapter, using traditional science involves potential competition. A researcher's payoff depends on the idea's maturation delay and on potential competition from the other researcher.

In the literature of preemption games, which are timing games with a first-mover advantage, a seminal paper is a model applied to technological competition, from Fudenberg and Tirole (1985). They proved that, for a duopoly, firms' equilibrium payoffs are equalized and rents are fully dissipated. This is not the case in my model since there is an asymmmetry of information between players, which leaves rents to researchers. Hendricks (1992) extended Fudenberg and Tirole's paper (1985) and showed that between being an innovator or an imitator, firms prefers to make believe they are imitators in order to soften the competition for the adoption of a new technology. This is linked to my setting because it is in the interest of the researchers to make believe they are not "born" yet. This is modeled by the breakthrough distribution, which is independent of researchers' will. In addition, Hopenhayn and Squintani (2010) studied preemption games applied to patent races, in which information is private. They showed that R&D secrecy slows down innovation disclosure compared to the public information case. In my model, I only study the private information case but it is easy to see that the public information case will lead to the same kind of results, that is, shorten the maturation delay before publishing.

Weeds (2002) and Lambrecht and Perraudin (2003) studied the case of real options. In Weeds' model (2002), firms who invest jointly in a project tend to delay their investment because of the patent-race which follows the breakthrough. Applied to preemption games, Lambrecht and Perraudin (2003) showed that firms face a tradeoff between learning about the future return of an uncertain project and the risk of being preempted when there is uncertainty about other's costs. On the contrary, in my model, players fear only potential competition and they do not know whether other players are working on the same idea.

More recently, Bobtcheff and Mariotti (2012) focused on uncertainty, which arises when a researcher who has experienced a breakthrough does not bring the others researchers up to date. Therefore they make their decisions relative to some probability distribution about the others' breakthrough times. The payoff function does not depend on depend on breakthrough time, whereas it is the case in my model. They found that in any Bayesian perfect equilibrium, distribution of players' moving times was the same. To compare with Hendricks' model (1992), in this paper, as everyone knows the breakthrough distribution, which is symmetric, there is "eroding reputation" whereas it does not exist in Hendricks' (1992). Brunnermeier and Morgan (2010) study a model in which players receive, at a random and secret times, signals about a payoff-relevant state variable. In this model, clocks are therefore not synchronized. They found that delay is decreased when clocks became more synchronized. By contrast, in my model, players clocks are synchronized.

The chapter is organized as follows. Section 2 describes the model. Section 3 studies the existence of pure-strategy equilibria. Section 4 concludes.

2 The Model

Time is continuous on $[0, +\infty)$. Time 0 represents the time of a pioneering discovery that opens up a new research field. Past that time, a new idea can be discovered. There are two identical scientists, 1 and 2. Calendar time is common knowledge, that is, they both know that at time 0, time starts to count. Both scientists are risk-neutral and discount future payoffs at the same rate. Scientist $i \in \{1, 2\}$ discovers the new idea at some random time $\tilde{\tau}^i$. After that time, they can publish the discovery. The two scientists fight for priority. It means that if one scientist publishes before the other, then the game stops and the other cannot publish. The scientists' discovery times $\tilde{\tau}^1$ and $\tilde{\tau}^2$ are independently drawn from an exponential distribution with parameter $\lambda > 0$. Accordingly, λ is called *the discovery speed*.

At her discovery time, a scientist (she) chooses between two technological choices: traditional science (TS) and citizen science (CS). With TS, she takes time to let the idea mature before publishing it. If a scientist born at τ chooses TS and publishes before her opponent at time $t \geq \tau$, then she obtains the publication payoff $L(t - \tau)$, which depends on the idea's maturation delay. If instead, she is preempted by her rival, she obtains 0. A scientist who discovers the idea at τ is called a *τ-born*. The two scientists cannot publish simultaneously. The payoff function $L : [0; +\infty) \to [0; +\infty)$ is continuous and twice continuously differentiable over $(0;+\infty)$. I make the following assumptions.

$$L(0) = 0 \text{ and } \dot{L}(0) = +\infty,$$
$$L(m) > 0 \text{ if } m > 0,$$
$$\dot{L}(m) > 0 \text{ if } m > M,$$
$$\dot{L}(m) < 0 \text{ if } m > M,$$
$$\ddot{L}(m) < 0 \text{ if } m < M.$$

In that way, L is positive, achieves his maximum in M and is increasing and concave between 0 and M. L represents for instance the discounted payoff of future rewards coming after the publication of the matured idea. Furthermore, $M > 0$ maximizes the scientist's TS publication payoff. M represents the optimal maturation delay that a scientist would choose if she was not threatened by preemption.

The second technological choice, CS, is not available anymore after the scientist's discovery time. Indeed, as she works by herself on the idea, she gets passionate into her work and becomes more reluctant to disclose information about her research. By choosing CS, she incurs a fixed cost $c \in (0, L(M))$ to publish instantaneously the matured discovery. This cost represents the investment to incur to create a platform through which citizens can participate. As soon as this platform is created, the high number of citizens who contribute enables the idea to be mature very quickly. However, as non-specialists are involved in the idea's maturation, the publication quality is uncertain. With probability $p \in (0, 1)$, the publication is of good quality and the scientist enjoys the optimal maturation payoff $L(M)$. With probability $1 - p$, the publication is of bad quality, and she enjoys no payoff. Besides, I make the following assumption.

Assumption 1.

$$c < pL(M)$$

In that way, the scientist's instantaneous payoff at her discovery time $pL(M) - c$ is strictly positive.[1] This ensures that there exists a trade-off between the two technological choices.

3 Equilibrium analysis

I focus on Bayesian pure-strategy equilibria. A scientist's strategy is composed by her technological choice and her publication time if she chooses TS. Consider the strategy in which the scientist chooses TS and publishes directly at the discovery time. Due to Assumption 1, she could always be better-off by choosing CS and enjoying the payoff $pL(M) - c$ instead of 0. Consequently, this strategy is strictly dominated and the scientist never chooses TS and publish at τ at equilibrium. In this way, a strategy for scientist $i \in \{1, 2\}$ is a function $\sigma^i : [0, \infty) \to [0, \infty)$ that specifies, for each possible value τ^i of her discovery time, the time $\sigma^i(\tau^i)$ at which she publishes her discovery. $\sigma^i(\tau^i) = \tau^i$ represents the equilibrium strategy of scientist i is she chooses CS. $\sigma^i(\tau^i) > \tau^i$ represents her equilibrium strategy if she publishes with TS at time $\sigma^i(\tau^i)$.

If scientist i chooses TS, her payoff depends on her type τ^i, on player j's strategy σ^j, and on her publication time t^i. It is equal to the publication payoff by letting the idea mature during time $t^i - \tau^i$, multiplied by the probability that her competitor will not publish before time t^i, knowing that he has not published before her discovery time τ^i:

$$\begin{aligned} V^i(t^i, \tau^i, \sigma^j) &= P(\sigma^j(\tilde{\tau}^j) > t^i | \sigma^j(\tilde{\tau}^j) > \tau^i) L(t^i - \tau^i) \\ &= \frac{P(\{\sigma^j(\tilde{\tau}^j) > t^i\} \cap \{\sigma^j(\tilde{\tau}^j) > \tau^i\})}{P(\sigma^j(\tilde{\tau}^j) > \tau^i)} L(t^i - \tau^i) \\ V^i(t^i, \tau^i, \sigma^j) &= \frac{P(\sigma^j(\tilde{\tau}^j) > t^i)}{P(\sigma^j(\tilde{\tau}^j) > \tau^i)} L(t^i - \tau^i) \end{aligned} \tag{1}$$

[1]Results still hold for any strictly positive payoff which is less than the optimally matured payoff.

The last equality comes from the fact that the publication time is always greater than the discovery time.

Definition 1. A pair (σ^1, σ^2) is a Bayesian equilibrium if for each i, $\tau^i \geq 0$ and $t^i \geq \tau^i$:

1. If scientist i publishes with TS at time $\sigma^i(\tau^i)$, her payoff must be at the same time greater than if she publishes with TS at any other time and greater than if she chooses CS:

$$V^i(\sigma^i(\tau^i), \tau^i, \sigma^j) \geq V^i(t^i, \tau^i, \sigma^j) \tag{2}$$

$$V^i(\sigma^i(\tau^i), \tau^i, \sigma^j) \geq pL(M) - c \tag{3}$$

2. If scientist i chooses CS, her payoff must be greater than if she publishes with TS at any time $t^i > \tau^i$, that is,

$$pL(M) - c \geq V^i(t^i, \tau^i, \sigma^j) \tag{4}$$

First, I focus on symmetric equilibria, that is, $\sigma^i(\tau^i) = \sigma(\tau)$ for all $i \in \{1, 2\}$. If scientists choose TS, then from a standard revealed-preference argument, using Inequality 2, the publication time is non decreasing in the discovery time.

Lemma 1. *If both scientists choose TS, then σ is weakly increasing in τ.*

In addition, I make the following assumption, which helps for the equilibrium analysis.

Assumption 2. If both scientists choose TS, σ is strictly increasing in τ.

In the following section, I focus on the existence of stationary equilibria, in which scientists choosing TS publish after a constant maturation delay, whatever their discovery time.

3.1 Existence of a symmetric stationary equilibrium

Intuitively, a possible equilibrium candidate would be the following: if one scientist is born before some threshold discovery time τ^*, she does not fear potential competition enough and let the idea mature a constant maturation delay (*cmd*). If instead she is born after that threshold, potential competition becomes too fierce and she chooses CS. A stationary equilibrium of this kind is depicted in Figure 1.

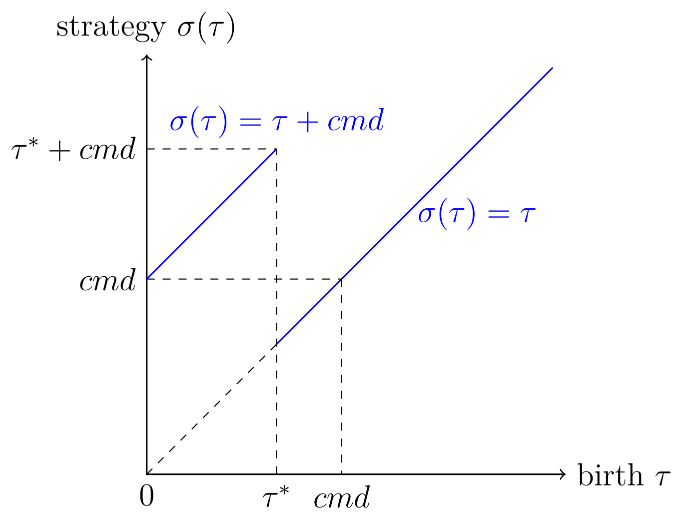

Figure 1 – a stationary equilibrium where scientists publish with TS after a constant maturation delay when born before τ^* and choose CS when born after τ^*

To prove the existence of an equilibrium of this kind, I must compare, at the scientist's discovery time, her ex-ante payoff if she publishes with TS after a constant maturation delay, to the CS payoff $pL(M) - c$. First, I study the scientists strategies if they choose TS. Assumption 2 implies that if scientists choose TS, the strategy σ can be inverted. In other words, $\phi = \sigma^{-1}$ is well defined and continuous on $\sigma([0, \infty))$. If a scientist chooses

TS, her maximization problem at her discovery time τ is the following:

$$\begin{aligned}\max_{t\geq\tau}\frac{P(\tilde{\sigma}(\tilde{\tau})>t)}{P(\tilde{\sigma}(\tilde{\tau})>\tau)}L(t-\tau) &= \max_{t\geq\tau}P(\tilde{\sigma}(\tilde{\tau})>t)L(t-\tau)\\ &= \max_{t\geq\tau}P(\tilde{\tau}>\phi(t))L(t-\tau)\\ &= \max_{t\geq\tau}e^{-\lambda.\phi(t)}L(t-\tau) \qquad (5)\end{aligned}$$

Taking the first order condition with respect to t, I find that the scientist publishes after a maturation delay M_λ which depends on the discovery speed λ but not on the discovery time τ. If she discovers the idea before the maturation delay M_λ, then the probability of preempt her competitor is the complementary of the probability that he discovers the idea between time 0 and time τ, which is $e^{-\lambda\tau}$. If she discovers the idea after time M_λ, then the probability of preempt her competitor is the complementary of the probability that he discovers the idea between time $\tau - M_\lambda$ and τ, which is $e^{-\lambda M_\lambda}$.

Lemma 2. *A scientist choosing TS publishes after a constant maturation delay M_λ and her payoff is equal to:*

$$\begin{cases} e^{-\lambda M_\lambda}L(M_\lambda) & \text{if} \quad \tau \geq M_\lambda \\ e^{-\lambda\tau}L(M_\lambda) & \text{if} \quad \tau < M_\lambda \end{cases}$$

The scientist's technological choice is determined by comparing the latter payoff to the CS payoff, $pL(M)-c$. When $pL(M)-c \geq L(M_\lambda)$, the 0-born scientist always prefers CS. It implies that any scientist discovering the idea after that time always prefers CS. Hence, there cannot exist a discontinuity in the scientists' equilibrium strategies. When $pL(M)-c < e^{-\lambda M_\lambda}L(M_\lambda)$, then no matter their discovery times, scientists always prefer the technological choice TS. In this way, there cannot exist a discontinuity in the equilibrium strategies.

Eventually, a discontinuous equilibrium potentially exists when

$$L(M_\lambda) \geq pL(M) - c > e^{-\lambda M_\lambda} L(M_\lambda)$$
$$\Leftrightarrow pL(M) - L(M_\lambda) \leq c < pL(M) - e^{-\lambda M_\lambda} L(M_\lambda)$$

This range of values is depicted in Figure 2. Denote $\tau^* \in [0, M_\lambda]$ the threshold discovery rate such that scientists born before that time publish with TS after the maturation delay M_λ and scientists born after choose CS. To simplify the analysis, I make the following assumption.

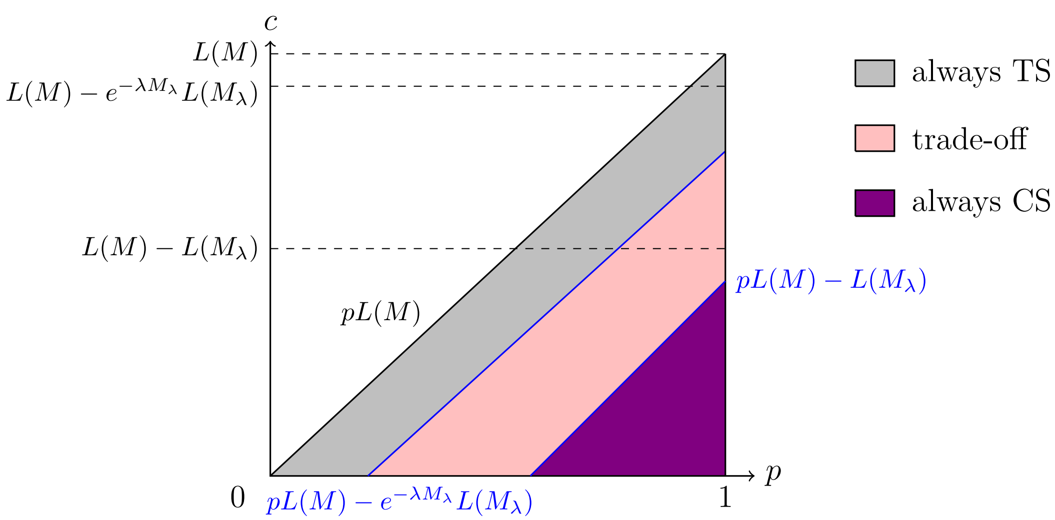

Figure 2 – a discontinuous stationary equilibrium may exists when the CS cost c is intermediate

Assumption 3.

$$\tau^* \leq M_\lambda$$

In this way, the two scientists always publish after time τ^*, that is, $P(\tilde{\sigma} > \tau^*) = 1$. A potential candidate for the threshold τ^* would be the discovery time at which the τ^*-born scientist is indifferent between the two technological choices. In other words, $e^{-\lambda \tau^*} L(M_\lambda) = pL(M) - c$. In fact, the scientist born at τ^* always strictly prefers CS.

Hence, it cannot be a threshold candidate. A second potential candidate would be such that $e^{-\lambda(\tau^{**}+M_\lambda)}L(M_\lambda) = e^{-\lambda\tau^*}L(M_\lambda)(= pL(M) - c)$. Then this candidate would be $\tau^{**} = \tau^* - M_\lambda < 0$, which is impossible. This leads to the following proposition.

Proposition 1. *There exists no symmetric stationary equilibrium in which there exists a $\tau^* \in [0; M_\lambda]$ such that*

- *Scientists born before τ^* publish with TS after the maturation delay M_λ;*
- *Scientists born after τ^* choose CS.*

As a matter of fact, the fear of being preempted is too high for early-born scientists so that they always have an interest in deviating by choosing CS. The technological choice TS is always left aside, which implies that the technological choice CS completely shifts the research landscape. Indeed, as soon as the latter choice is available for scientists, there is no adjustment delay during which they can adapt their research habits to the new CS technological change. This harms the early-borns scientists who are the first ones to adapt very quickly to the technological change. This profitable deviation implies that in the pink area in Figure 2, scientists always choose CS. Consequently, there exist only two kinds of stationary equilibria, which contain no discontinuity in the scientists' strategies.

Proposition 2. *There exist two kinds of symmetric stationary equilibria:*

- *If $c > pL(M) - e^{-\lambda M_\lambda}L(M_\lambda)$, scientists publish with TS after the maturation delay M_λ;*
- *If $c \leq pL(M) - e^{-\lambda M_\lambda}L(M_\lambda)$, scientists choose CS.*

If CS is expensive, scientists choose TS. Otherwise, they choose CS. To compare with BBM, where the technological choice CS does not exist, the equilibrium payoff remains the same in the first case. Consequently, scientists do not benefit from the existence of the CS technological choice. In the first case, without the existence of the technological choice CS, in BBM, a τ-born scientist born after the time corresponding to the equilibrium maturation delay M_λ enjoys the payoff $e^{-\lambda M_\lambda}L(M_\lambda)$. If she is born before that time, she enjoys the payoff $e^{-\lambda\tau}L(M_\lambda)$. This payoff must be compared to $pL(M) - c$, which is the payoff that every scientist enjoys with CS at equilibrium.

If $c > pL(M) - e^{-\lambda\tau}L(M_\lambda)$, the τ-born scientist is born early. Without the existence of the CS technological choice, she faces a low preemption risk and enjoys her position of discovering the idea early. She looses this advantage when every scientist chooses CS. Hence, her payoff is strictly lower than without the existence of the CS technological choice and she looses from CS. If instead, $e^{-\lambda\tau}L(M_\lambda) \leq pL(M) - c$, the τ-born scientist is born late. Without the existence of the CS technological choice, there is a high preemption risk, which is reflected in her low payoff. This risk vanishes when all scientists choose CS. Consequently, her payoff is strictly higher with CS and she benefits from the existence of the CS technological choice. Therefore, CS has an heterogeneous effect on scientists: it favors the late borns and harms the early borns.

To go further, the early-borns scientists who have the most interest to deviate from TS are the ones born just before the cut-off τ^*. Hence, if their equilibrium maturation delay becomes smaller when they discovers the idea closer to τ^*, their fear of being preempted is smaller. In this way, a discontinuous equilibrium may exist, in which the equilibrium scientists' strategy is non-stationary. In the following subsection, I study this kind of equilibrium.

3.2 Existence of a non-stationary discontinuous equilibrium

Suppose that there exists a symmetric equilibrium exhibiting a $\bar{\tau} \in (0, +\infty)$ such that every scientist born before $\bar{\tau}$ chooses TS and everybody after chooses CS. To simplify the analysis, I make the following assumption.

Assumption 4.

$$\sigma(0) > \bar{\tau}.$$

It ensures that all scientists publish after time $\bar{\tau}$, that is, $P(\tilde{\sigma} > \bar{\tau}) = 1$. A discontinuous equilibrium satisfying this assumption is depicted in Figure 3. At equilibrium, a scientist born just before $\bar{\tau}$ chooses TS and waits whereas another one born just after $\bar{\tau}$ chooses CS. As a matter of fact, the difference between the equilibrium payoff of these two scientists tends to 0 as the discovery time of the former tends to $\bar{\tau}$. It implies that there cannot exist a discontinuity in the equilibrium payoff function at $\bar{\tau}$. This leads to the following lemma.

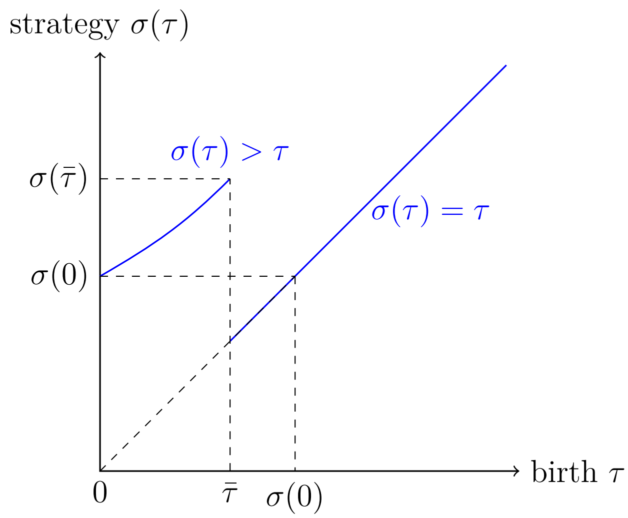

Figure 3 – a discontinuous equilibrium satisfying Assumption 4, where scientists choose TS when born before $\bar{\tau}$ and choose CS when born after

Lemma 3. *The equilibrium payoff function is continuous on* $[0; +\infty]$.

In that way, the payoff of the $\bar{\tau}$-born scientist remains the same with the two technological choices. Consequently, she must be indifferent between the two options.

Hence, for $\epsilon > 0$ small enough, I have

$$\frac{P(\tilde{\sigma} > \sigma(\bar{\tau} - \epsilon))}{P(\tilde{\sigma} > \bar{\tau} - \epsilon)} L(\sigma(\bar{\tau} - \epsilon) - (\bar{\tau} - \epsilon)) \geq \frac{P(\tilde{\sigma} > \sigma(\bar{\tau}))}{P(\tilde{\sigma} > \bar{\tau})} L(\sigma(\bar{\tau}) - \bar{\tau}).$$

It means that, if ϵ is small enough, the $(\bar{\tau} - \epsilon)$-born scientist has a larger equilibrium payoff than the $\bar{\tau}$-born one. By solving the former's maximization problem, knowing that the latter is indifferent between the two technological choices, she must have discovered the idea before time 0. It follows that, necessarily, at equilibrium, the $\bar{\tau}$-born scientist always strictly prefers CS to TS. Therefore, such a discontinuous equilibrium does not exist. The essence of the proof is similar to Proposition 1.

Proposition 3. *There exists no symmetric equilibrium in which there exists a $\bar{\tau}$ in $[0; \sigma(0))$ such that scientists born before $\bar{\tau}$ choose TS and the ones born after or at $\bar{\tau}$ choose CS.*

As in Proposition 1, the non-existence of this type of equilibrium implies that the early-borns scientists must adapt quickly to the technological change CS. Even scientists discovering the idea just before $\bar{\tau}$, and publishing after a small maturation delay, have interest in deviating by choosing CS. Besides, from Lemma 3, the $\bar{\tau}$-born scientist is indifferent between the two technological choices. In this way, an equilibrium containing an equilibrium strategy which is continuous at $\bar{\tau}$ may exist. This is what I study in the next subsection.

3.3 Existence of a non-stationary continuous equilibrium

Suppose that there exists a symmetric equilibrium with a $\bar{\tau} \in (0, +\infty)$ such that scientists born before $\bar{\tau}$ choose TS and the ones born after choose CS. I assume that σ is continuous on $[0, +\infty)$. An equilibrium of this type is depicted in Figure 4.

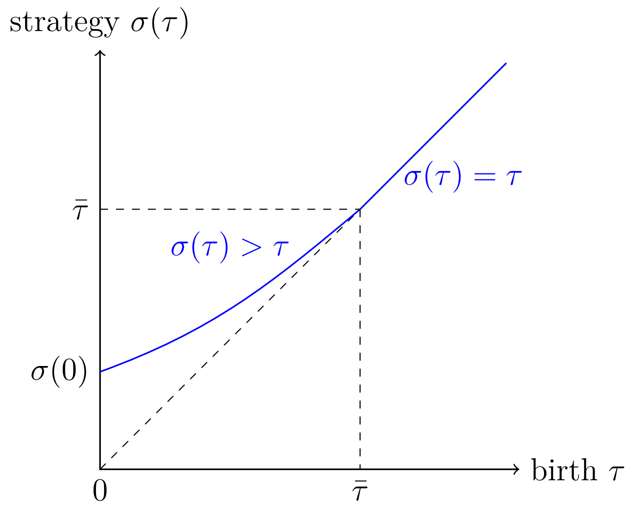

Figure 4 – a continuous equilibrium in which scientists choose TS when born before $\bar{\tau}$ and choose CS when born after

Because σ is left-continuous at $\bar{\tau}$, the equilibrium maturation delay $\sigma(\tau) - \tau$ is decreasing in τ over $[0; \bar{\tau})$. In this way, the equilibrium payoff function tends to 0 on the left of $\bar{\tau}$ as τ tends to $\bar{\tau}$. Besides, the $\bar{\tau}$-born scientist chooses CS. It implies that if the discovery time is close enough below $\bar{\tau}$, the equilibrium payoff of a scientist born at that time who chooses TS is lower than the CS payoff. Hence, she strictly prefers CS to TS. Here lies a contradiction and there exists no such equilibrium.

Proposition 4. *There exists no symmetric equilibrium in which there exists a* $\bar{\tau} \in [0; +\infty)$ *such that:*

- σ *is continuous on* $[0, +\infty)$*;*
- *Scientists born before* $\bar{\tau}$ *choose TS;*
- *Scientists born after or at* $\bar{\tau}$ *choose CS.*

In this kind of equilibrium, early-born scientists would have shifted softly from TS to CS. Unfortunately, the latter does not exist. Together with Propositions 1 and 3, it means that if CS is low-cost, scientists must adapt quickly to the technological change.

Furthermore, the latter creates a congestion in the use of CS. This technological choice requires the help of citizens, who have limited attention and time. In this way, they cannot work for both scientists at the same time. This questions the existence of an asymmetric equilibrium, in which the two scientists would have different technological choices. This is what I study in the following subsection.

3.4 Existence of an asymmetric stationary equilibrium

Suppose that, at equilibrium, scientist i chooses CS and scientist j publishes with TS after a maturation delay M_λ. This equilibrium must satisfy Inequality 4 for scientist i and Inequalities 2 and 3 for scientist j, that is, for any discovery times $\tau^i \geq 0$ and $\tau^j \geq 0$, and for any publication times $t^i \geq \tau^i$ and $t^j \geq \tau^j$,

$$\begin{cases} pL(M) - c & \geq V^i(t^i, \tau^i, \tau^j + M_\lambda) \\ V^j(\tau^j + M_\lambda, \tau^j, \tau^i) & \geq V^j(t^j, \tau^j, \tau^i) \\ V^j(\tau^j + M_\lambda, \tau^j, \tau^i) & \geq pL(M) - c \end{cases} \tag{6}$$

Inequality 6 implies that the CS cost belongs to the grey area in Figure 2. As a matter of fact, a scientist who discovers the idea before the maturation delay M_λ has no preemption risk in a symmetric equilibrium. By contrast, in such an asymmetric equilibrium, there exists a preemption risk that her comptetitor publishing instantaneously with CS discovers the idea before her. This preemption risk is high so that she has a strictly profitable deviation by choosing CS. In this way, there exists no asymmetric equilibrium in which one scientist chooses CS and the other publishes with TS after waiting a delay M_λ.

Proposition 5. *There exists no asymmetric equilibrium in which one scientist chooses CS and the other publishes with TS after the maturation delay M_λ.*

This type of equilibrium would have eliminated the congestion problem with CS. In addition, in terms of policy implications, such an equilibrium would have benefited the government fundings. Indeed, if money has to be allocated to citizen science, then it must be splitted between the two scientists.

4 Conclusion

In this chapter, I studied the scientists' trade-off between citizen science and traditional science when they fight for preemption. I prove that there exist no equilibrium such that scientists born early choose TS and the ones born late choose CS. There always exists a profitable deviation for a scientist born close below the discovery time threshold which is to choose CS. This harms the early-born scientists who have no time to adapt to the technological change. Interestingly, there exist two kinds of symmetric stationnary equilibria: if CS is a low-cost technological choice, then every scientist chooses CS. Otherwise, everyone chooses TS and waits a constant maturation delay before publication. Researchers in the second kind receive the same payoff as BBM, whereas in the first kind, early borns are harmed and late borns are better-off than without the existence of CS. Hence, my model increases preemption fear compare to BBM.

This chapter provides some government policies to adopt regarding CS. It depends on whether the government (he) is interested in maximizing the scientists' payoff or letting the idea mature optimally. If his objective is the second, he should always encourage CS as the idea is always mature at its best whereas the maturation delay is stricly lower with TS. If the government is interested by maximizing the scientists' payoff, then his action has no impact if CS is expensive. If CS is low-cost, then everybody choosing CS at equilibrium harms the early-borns and benefits the late-borns. If the government aims at increasing the publication speed, then he should encourage CS. If he puts more weight on the early-borns than on the late-borns, then he should restrict the CS use.

To go further, it would be interesting to prove is that the equilibrium publication time increasing in the discovery time is a necessary condition to have an equilibrium. In this way, similarly to Proposition 4, I would prove that there cannot exist a discontinuity point in the equilibrium publication time. Hence the only existing symmetric equilibrium would be that every scientist chooses CS.

In a more general setting, scientists could pay a cost in a continuous range of values to reduce the maturation delay. In this way, they could enjoy the same publication payoff than the model without investment but earlier in time. This modelization reflects well the scientist's allocation of time between several ideas. Should she invest more time in ideas which brings a higher publication payoff later or a lower payoff sooner? This setting would increase the concavity of the payoff function and the optimal maturation payoff will be achieved earlier. At equilibrium, they will choose both the publication time and the investment cost. The equilibrium analysis is not obvious since the scientist's choice is two-dimensional.

5 Appendix

Proof of Lemma 1

Procede by contradiction. Suppose that there exist τ and $\hat{\tau}$ in $[0;\bar{\tau}]$ such that $\hat{\tau} \geq \tau$ and $\sigma(\tau) > \sigma(\hat{\tau})$. As τ and $\hat{\tau}$ both satisfy Inequality 2, then

$$\frac{P(\tilde{\sigma} > \sigma(\tau))}{P(\tilde{\sigma} > \tau)} L(\sigma(\tau) - \tau) \geq \frac{P(\tilde{\sigma} > \sigma(\hat{\tau}))}{P(\tilde{\sigma} > \tau)} L(\sigma(\hat{\tau}) - \tau)$$
$$\frac{P(\tilde{\sigma} > \sigma(\hat{\tau}))}{P(\tilde{\sigma} > \hat{\tau})} L(\sigma(\hat{\tau}) - \hat{\tau}) \geq \frac{P(\tilde{\sigma} > \sigma(\tau))}{P(\tilde{\sigma} > \hat{\tau})} L(\sigma(\tau) - \hat{\tau})$$

By adding the two latter inequalities, I have

$$P(\tilde{\sigma} > \sigma(\tau))L(\sigma(\tau) - \tau) + P(\tilde{\sigma} > \sigma(\hat{\tau}))L(\sigma(\hat{\tau}) - \hat{\tau}) \geq P(\tilde{\sigma} > \sigma(\hat{\tau}))L(\sigma(\hat{\tau}) - \tau) + P(\tilde{\sigma} > \sigma(\tau))L(\sigma(\tau) - \hat{\tau})$$
$$\Leftrightarrow P(\tilde{\sigma} > \sigma(\tau))(L(\sigma(\tau) - \tau) - L(\sigma(\tau) - \hat{\tau})) \geq P(\tilde{\sigma} > \sigma(\hat{\tau}))(L(\sigma(\hat{\tau}) - \tau) - L(\sigma(\hat{\tau}) - \hat{\tau})) \quad (7)$$

L is increasing and concave on $[0, M]$, hence

$$0 < L(\sigma(\tau) - \tau) - L(\sigma(\tau) - \hat{\tau}) < L(\sigma(\hat{\tau}) - \tau) - L(\sigma(\hat{\tau}) - \hat{\tau}) \quad (8)$$

Moreover, $P(\tilde{\sigma} > .)$ is non increasing, which implies that

$$P(\tilde{\sigma} > \sigma(\tau)) \leq P(\tilde{\sigma} > \sigma(\hat{\tau})) \quad (9)$$

But from Inequalities 7 and 8, I must have

$$P(\tilde{\sigma} > \sigma(\tau)) > P(\tilde{\sigma} > \sigma(\hat{\tau}))$$

This contradicts Inequality 9. Consequently, σ is nondecreasing in τ.

Proof of Lemma 2

The first order condition with respect to t is:

$$\begin{aligned}
-\lambda.\dot{\phi}(t)e^{-\lambda.\phi(t)}L(t-\phi(t)) + e^{-\lambda.\phi(t)}\dot{L}(t-\phi(t)) = 0 \\
\Leftrightarrow \lambda.\dot{\phi}(t) = \frac{\dot{L}(t-\phi(t))}{L(t-\phi(t))} \\
\Leftrightarrow \dot{\phi}(t) = \frac{1}{\lambda}\frac{\dot{L}}{L}(t-\phi(t))
\end{aligned}$$

A solution to this EDO is

$$\phi_\lambda(t) = t - \Big(\frac{\dot{L}}{L}\Big)^{-1}(\lambda)$$

It implies that the maturation delay, M_λ, depends on the discovery rate λ but not on the discovery time τ:

$$M_\lambda = t - \phi_\lambda(t) = \Big(\frac{\dot{L}}{L}\Big)^{-1}(\lambda)$$

Besides,

$$\begin{aligned}
P(\tilde{\tau} > \phi(\tau)) &= e^{-\lambda\phi(\tau)} \\
&= \begin{cases} e^{-\lambda(\tau - M_\lambda)} & \text{if} \qquad \tau > M_\lambda \\ 1 & \text{otherwise} \end{cases} \\
&= e^{-\lambda max\{0;\tau - M_\lambda\}}
\end{aligned}$$

Therefore, the ex-ante payoff of a τ-born scientist who chooses TS, described in

Equation 5, becomes

$$\frac{P(\tilde{\tau} > \phi(\tau + M_\lambda))}{P(\tilde{\tau} > \phi(\tau))} L(M_\lambda) = \frac{e^{-\lambda\tau}}{e^{-\lambda max\{0;\tau - M_\lambda\}}} L(M_\lambda)$$
$$= \begin{cases} e^{-\lambda M_\lambda} L(M_\lambda) & \text{if} \quad \tau \geq M_\lambda \\ e^{-\lambda\tau} L(M_\lambda) & \text{if} \quad \tau < M_\lambda \end{cases}$$

Proof of Proposition 1

I look for a profitable deviation for the τ^*-born scientist. Her payoff when choosing TS and waiting M_λ before publication is $\frac{P(\tilde{\sigma} > \tau^* + M_\lambda)}{P(\tilde{\sigma} > \tau^*)} L(M_\lambda)$.
After $\tau^* + M_\lambda$, the only scientists who did not publish yet are born after τ^*, that is,

$$P(\tilde{\sigma} > \tau^* + M_\lambda) = P(\tilde{\tau} > \tau^* + M_\lambda).$$

Moreover, due to Assumption 3, $P(\tilde{\sigma} > \tau^*) = 1$. Then

$$\frac{P(\tilde{\sigma} > \tau^* + M_\lambda)}{P(\tilde{\sigma} > \tau^*)} L(M_\lambda) = P(\tilde{\tau} > \tau^* + M_\lambda) L(M_\lambda) = e^{-\lambda(\tau^* + M_\lambda)} L(M_\lambda).$$

By definition, τ^* is such that $e^{-\lambda\tau^*} L(M_\lambda) = pL(M) - c$. It implies that

$$\frac{P(\tilde{\sigma} > \tau^* + M_\lambda)}{P(\tilde{\sigma} > \tau^*)} L(M_\lambda) = e^{-\lambda(\tau^* + M_\lambda)} L(M_\lambda) = (pL(M) - c)e^{-\lambda M_\lambda} < pL(M) - c.$$

Therefore, the τ^*-born has a profitable deviation which to choose CS. In this way, a discontinuous equilibrium with such a threshold τ^* does not exist. A second potential candidate would be τ^{**} such that

$$e^{-\lambda(\tau^{**} + M_\lambda)} L(M_\lambda)(= pL(M) - c) = e^{-\lambda\tau^*} L(M_\lambda),$$

but then $\tau^{**} = \tau^* - M_\lambda < 0$, which is impossible. Consequently, there exists no such τ^{**} in $[0; M_\lambda]$. For any τ in $[0; M_\lambda]$, it is always profitable to choose CS instead of waiting M_λ with TS. Hence, a discontinuous stationnary equilibrium does not exist.

Proof of Proposition 2

Assume that $pL(M) - c \geq e^{-\lambda M_\lambda L(M_\lambda)}$ and fix $\tau \geq 0$. Suppose that both scientists choose CS and look for a profitable deviation for the τ-born scientist. The strategy of her opponent is $\tilde{\sigma}(\tilde{\tau}) = \tilde{\tau}$. If the τ-born scientist deviates by choosing TS and publishing at time $t > \tau$, she obtains the following payoff:

$$\frac{P(\tilde{\sigma} > t)}{P(\tilde{\sigma} > \tau)} L(t - \tau) = \frac{P(\tilde{\tau} > t)}{P(\tilde{\tau} > \tau)} L(t - \tau) = e^{-\lambda(t-\tau)} L(t - \tau)$$

Her maximization problem is

$$\begin{aligned} &\max_t \{e^{-\lambda(t-\tau)} L(t - \tau)\} \\ &\text{s.t. } e^{-\lambda(t-\tau)} L(t - \tau) \geq pL(M) - c \end{aligned}$$

The FOC of the unconstrained problem is:

$$\begin{aligned} & -\lambda e^{-\lambda(t-\tau)} L(t - \tau) + e^{-\lambda(t-\tau)} \dot{L}(t - \tau) = 0 \\ \Leftrightarrow & \frac{\dot{L}}{L}(t - \tau) = \lambda \\ \Leftrightarrow & t = \tau + (\frac{\dot{L}}{L})^{-1}(\lambda) = \tau + M_\lambda \end{aligned} \tag{10}$$

Then the scientist has a payoff of $e^{-\lambda M_\lambda} L(M_\lambda)$. The constrained problem payoff would be even lower. However, $pL(M) - c \geq e^{-\lambda M_\lambda} L(M_\lambda)$, therefore the scientist has no profitable deviation. This result holds for any τ. Consequently, the stationnary equilibrium in which scientists always choose CS is sustainable.

Now assume that $pL(M) - c < e^{-\lambda M_\lambda} L(M_\lambda)$ and fix $\tau \geq 0$. Suppose that both scientists choose TS and wait M_λ before publication. A τ-born scientist enjoys the payoff

$$\begin{cases} e^{-\lambda M_\lambda} L(M_\lambda) & \text{if } \tau \geq M_\lambda, \\ e^{-\lambda \tau} L(M_\lambda) & \text{if } \tau < M_\lambda. \end{cases}$$

Besides, for all $\tau \in [0; M_\lambda]$,

$$e^{-\lambda \tau} L(M_\lambda) \geq e^{-\lambda M_\lambda} L(M_\lambda) > pL(M) - c.$$

Therefore there is no profitable deviation for the τ-born scientist and the stationnary equilibrium in which scientists choose TS and wait M_λ before publication is sustainable.

Proof of Lemma 3

Over the interval $[0, \bar{\tau}]$, the equilibrium payoff function is equal to $\frac{P(\tilde{\sigma} > \sigma(\tau))}{P(\tilde{\sigma} > \tau)} L(\sigma(\tau) - \tau)$, which is continuous in τ. Over the interval $(\bar{\tau}, +\infty)$, the equilibrium payoff function is equal to $pL(M) - c$, which is continuous in τ. The remaining point is to prove that the equilibrium payoff function is continuous at $\bar{\tau}$. Fix $\epsilon > 0$ and look at the $(\bar{\tau} - \epsilon)$-born scientist. Her equilibrium payoff should tend to $pL(M) - c$ as ϵ tends to 0. At equilibrium, as she chooses TS and wait until $\bar{\tau} - \epsilon$ before publication, her payoff should be greater than if she chooses CS, that is,

$$\frac{P(\tilde{\sigma} > \sigma(\bar{\tau} - \epsilon))}{P(\tilde{\sigma} > \bar{\tau} - \epsilon)} L(\sigma(\bar{\tau} - \epsilon) - (\bar{\tau} - \epsilon)) - pL(M) - c \geq 0$$

Besides, at equilibrium, the $(\bar{\tau} + \epsilon)$-born scientist prefers CS than TS and publishing at $\sigma(\bar{\tau})$, that is,

$$pL(M) - c \geq \frac{P(\tilde{\sigma} > \sigma(\bar{\tau}))}{P(\tilde{\sigma} > \bar{\tau} + \epsilon)} L(\sigma(\bar{\tau}) - (\bar{\tau} + \epsilon))$$

It implies that

$$\begin{aligned} &\frac{P(\tilde{\sigma} > \sigma(\bar{\tau} - \epsilon))}{P(\tilde{\sigma} > \bar{\tau} - \epsilon)} L(\sigma(\bar{\tau} - \epsilon) - (\bar{\tau} - \epsilon)) - \frac{P(\tilde{\sigma} > \sigma(\bar{\tau}))}{P(\tilde{\sigma} > \bar{\tau} + \epsilon)} L(\sigma(\bar{\tau}) - (\bar{\tau} + \epsilon)) \\ &\qquad \geq \frac{P(\tilde{\sigma} > \sigma(\bar{\tau} - \epsilon))}{P(\tilde{\sigma} > \bar{\tau} - \epsilon)} L(\sigma(\bar{\tau} - \epsilon) - (\bar{\tau} - \epsilon)) - pL(M) - c \geq 0 \end{aligned}$$

Therefore,

$$\begin{aligned} &\left|\frac{P(\tilde{\sigma} > \sigma(\bar{\tau} - \epsilon))}{P(\tilde{\sigma} > \bar{\tau} - \epsilon)} L(\sigma(\bar{\tau} - \epsilon) - (\bar{\tau} - \epsilon)) - pL(M) - c\right| \\ &\qquad \leq \left|\frac{P(\tilde{\sigma} > \sigma(\bar{\tau} - \epsilon))}{P(\tilde{\sigma} > \bar{\tau} - \epsilon)} L(\sigma(\bar{\tau} - \epsilon) - (\bar{\tau} - \epsilon)) - \frac{P(\tilde{\sigma} > \sigma(\bar{\tau}))}{P(\tilde{\sigma} > \bar{\tau} + \epsilon)} L(\sigma(\bar{\tau}) - (\bar{\tau} + \epsilon))\right| \\ &\qquad \leq \left|\frac{P(\tilde{\sigma} > \sigma(\bar{\tau} - \epsilon))}{P(\tilde{\sigma} > \bar{\tau} - \epsilon)} L(\sigma(\bar{\tau} - \epsilon) - (\bar{\tau} - \epsilon)) - \frac{P(\tilde{\sigma} > \sigma(\bar{\tau}))}{P(\tilde{\sigma} > \bar{\tau})} L(\sigma(\bar{\tau}) - \bar{\tau})\right| \\ &\qquad\quad + \left|\frac{P(\tilde{\sigma} > \sigma(\bar{\tau}))}{P(\tilde{\sigma} > \bar{\tau})} L(\sigma(\bar{\tau}) - \bar{\tau}) - \frac{P(\tilde{\sigma} > \sigma(\bar{\tau}))}{P(\tilde{\sigma} > \bar{\tau} + \epsilon)} L(\sigma(\bar{\tau}) - (\bar{\tau} + \epsilon))\right| \end{aligned}$$

The equilibrium payoff function is left-continuous at $\bar{\tau}$. Hence, the first part of the right-hand side of the inequality tends to 0 as ϵ tends to 0. The second part of the right-hand side of the inequality is equal to

$$P(\tilde{\sigma} > \sigma(\bar{\tau})) \left(\frac{L(\sigma(\bar{\tau}) - \bar{\tau})}{P(\tilde{\sigma} > \bar{\tau})} - \frac{L(\sigma(\bar{\tau}) - (\bar{\tau} + \epsilon))}{P(\tilde{\sigma} > \bar{\tau} + \epsilon)} \right)$$

L is continuous on $[0, +\infty)$, then $\lim_{\epsilon \to 0} L(\sigma(\bar{\tau}) - (\bar{\tau} + \epsilon)) = L(\sigma(\bar{\tau}) - \bar{\tau})$. Moreover, for ϵ small enough,

$$\begin{aligned} P(\tilde{\sigma} > \bar{\tau} + \epsilon) &= P(\tilde{\tau} < \bar{\tau}) + P(\tilde{\tau} > \bar{\tau} + \epsilon) \\ &= 1 - e^{-\lambda\bar{\tau}} + e^{-\lambda(\bar{\tau} + \epsilon)} < 1 \\ \Rightarrow \lim_{\epsilon \to 0} P(\tilde{\sigma} > \bar{\tau} + \epsilon) &= 1 - e^{-\lambda\bar{\tau}} + e^{-\lambda\bar{\tau}} = 1 \end{aligned}$$

Therefore
$$\lim_{\epsilon \to 0} \left(\frac{L(\sigma(\bar{\tau}) - \bar{\tau})}{P(\tilde{\sigma} > \bar{\tau})} - \frac{L(\sigma(\bar{\tau}) - (\bar{\tau} + \epsilon))}{P(\tilde{\sigma} > \bar{\tau} + \epsilon)} \right) = 0$$

It implies that

$$0 \leq | \frac{P(\tilde{\sigma} > \sigma(\bar{\tau} - \epsilon))}{P(\tilde{\sigma} > \bar{\tau} - \epsilon)} L(\sigma(\bar{\tau} - \epsilon) - (\bar{\tau} - \epsilon)) - pL(M) - c| \leq \xi(\epsilon)$$

with $\xi(\epsilon) \geq 0$ and $\lim_{\epsilon \to 0} \xi(\epsilon) = 0$. Therefore, by the Sandwich theorem,

$$\lim_{\epsilon \to 0} \frac{P(\tilde{\sigma} > \sigma(\bar{\tau} - \epsilon))}{P(\tilde{\sigma} > \bar{\tau} - \epsilon)} L(\sigma(\bar{\tau} - \epsilon) - (\bar{\tau} - \epsilon)) = pL(M) - c$$
$$\text{and } \frac{P(\tilde{\sigma} > \sigma(\bar{\tau}))}{P(\tilde{\sigma} > \bar{\tau})} L(\sigma(\bar{\tau}) - \bar{\tau}) = pL(M) - c$$

Consequently, the equilibrium payoff function is continuous at $\bar{\tau}$, hence, continuous on $[0; +\infty)$.

Proof of Proposition 3

With Lemma 3, I have proved that the $\bar{\tau}$-born scientist is indifferent between the two technological choices, that is,

$$\frac{P(\tilde{\sigma} > \sigma(\bar{\tau}))}{P(\tilde{\sigma} > \bar{\tau})} L(\sigma(\bar{\tau}) - \bar{\tau}) = pL(M) - c. \tag{11}$$

Besides,

$$P(\tilde{\sigma} > \sigma(\bar{\tau})) = P(\tilde{\tau} > \sigma(\bar{\tau})) = e^{-\lambda.\sigma(\bar{\tau})}$$

Therefore, Equation 11 becomes

$$e^{-\lambda.\sigma(\bar{\tau})} L(\sigma(\bar{\tau}) - \bar{\tau}) = pL(M) - c \tag{12}$$

Every scientist born after $\bar{\tau}$ chooses CS, hence, $P(\tilde{\sigma} > t) = P(\tilde{\tau} > t) = e^{-\lambda t}$. The best strategy $\sigma(\bar{\tau})$ of the $\bar{\tau}$-born scientist satisfies the FOC to her maximization problem, that is,

$$\max_{t \geq \bar{\tau}} \frac{P(\tilde{\sigma} > t)}{P(\tilde{\sigma} > \bar{\tau})} L(t - \bar{\tau}) = \max_{t \geq \bar{\tau}} P(\tilde{\sigma} > t) L(t - \bar{\tau}) = \max_{t \geq \bar{\tau}} e^{-\lambda t} L(t - \bar{\tau})$$

The FOC is

$$\begin{aligned} & -\lambda e^{-\lambda \sigma(\bar{\tau})} L(\sigma(\bar{\tau}) - \bar{\tau}) + e^{-\lambda \sigma(\bar{\tau})} \dot{L}(\sigma(\bar{\tau}) - \bar{\tau}) = 0 \\ \Leftrightarrow & \frac{\dot{L}}{L}(\sigma(\bar{\tau}) - \bar{\tau}) = \lambda \\ \Leftrightarrow & \sigma(\bar{\tau}) = \bar{\tau} + (\frac{\dot{L}}{L})^{-1}(\lambda) = \bar{\tau} + M_\lambda \end{aligned}$$

Replacing in Equation 12, I have

$$e^{-\lambda(\bar{\tau} + M_\lambda)} L(M_\lambda) = pL(M) - c$$

This means that $\bar{\tau} = \tau^{**}$, which is the threshold found in Proposition 1. However, $\tau^{**} < 0$, which implies that there exists no such $\bar{\tau}$ in $[0; +\infty)$. Consequently, there exists no discontinuous equilibrium with a threshold $\bar{\tau}$ in $[0, \sigma(0)]$ below which scientists born before that threshold choose TS and above which the ones born after choose CS.

Proof of Proposition 4

Fix $\epsilon > 0$. At equilibrium, the $(\bar{\tau} - \epsilon)$-born scientist enjoys from TS the following payoff:

$$\frac{P(\tilde{\sigma} > \sigma(\bar{\tau} - \epsilon))}{P(\tilde{\sigma} > \bar{\tau} - \epsilon)} L(\sigma(\bar{\tau} - \epsilon) - (\bar{\tau} - \epsilon))$$

As $\sigma(\bar{\tau}) = \bar{\tau}$,

$$\lim_{\epsilon \to 0} \frac{P(\tilde{\sigma} > \sigma(\bar{\tau} - \epsilon))}{P(\tilde{\sigma} > \bar{\tau} - \epsilon)} = \frac{P(\tilde{\sigma} > \sigma(\bar{\tau}))}{P(\tilde{\sigma} > \bar{\tau})} = 1$$

Besides,

$$\begin{aligned} &\lim_{\epsilon \to 0} \sigma(\bar{\tau} - \epsilon) - (\bar{\tau} - \epsilon) = \sigma(\bar{\tau}) - \bar{\tau} = 0 \\ \Rightarrow &\lim_{\epsilon \to 0} L(\sigma(\bar{\tau} - \epsilon) - (\bar{\tau} - \epsilon)) = L(0) = 0 \end{aligned}$$

It implies that, as ϵ tends to 0, the equilibrium payoff of the $\bar{\tau}$-born scientist tends to 0, which is smaller than $pL(M) - c$. Hence, she strictly prefers CS than TS. This holds for any $\bar{\tau}$ in $[0; +\infty)$. Consequently, there exist no such a discontinuous equilibrium.

Proof of Proposition 5

From Equality 1, I have

$$\begin{aligned} V^i(t^i, \tau^i, \tau^j + M_\lambda) &= \frac{P(\tilde{\tau}^j + M_\lambda > t^i)}{P(\tilde{\tau}^j + M_\lambda > \tau^i)} L(t^i - \tau^i) \\ V^j(\tau^j + M_\lambda, \tau^j, \tau^i) &= \frac{P(\tilde{\tau}^i > \tau^j + M_\lambda)}{P(\tilde{\tau}^i > \tau^j)} L(M_\lambda) \\ &= e^{-\lambda M_\lambda} L(M_\lambda) \\ V^j(t^j, \tau^j, \tau^i) &= \frac{P(\tilde{\tau}^i > t^j)}{P(\tilde{\tau}^i > \tau^j)} L(t^j - \tau^j) \\ &= e^{-\lambda(t^j - \tau^j)} L(t^j - \tau^j) \end{aligned}$$

The solution to the maximization problem

$$\max_{t^j} e^{-\lambda(t^j - \tau^j)} L(t^j - \tau^j)$$

is given by Equation 10. Hence, the optimal maturation delay is M_λ. Besides,

$$\begin{aligned}\frac{P(\tilde{\tau}^j + M_\lambda > t^i)}{P(\tilde{\tau}^j + M_\lambda > \tau^i)} &= \frac{P(\tilde{\tau}^j > t^i - M_\lambda)}{P(\tilde{\tau}^j > \tau^i - M_\lambda)} \\ &= \frac{e^{-\lambda \max\{0, t^i - M_\lambda\}}}{e^{-\lambda \max\{0, \tau^i - M_\lambda\}}} \\ &= \begin{cases} e^{-\lambda(t^i - \tau^i)} & \text{if } M_\lambda \leq \tau^i \\ e^{-\lambda(t^i - M_\lambda)} & \text{if } \tau^i < M_\lambda \leq t^i \\ 1 & \text{if } M_\lambda > t^i \end{cases}\end{aligned}$$

There are three cases.

1) If $\tau^i \geq M_\lambda$, the conditions for this equilibrium to exist become

$$\begin{cases} e^{-\lambda M_\lambda} L(M_\lambda) & \geq pL(M) - c \geq e^{-\lambda(t^i - \tau^i)} L(t^i - \tau^i) \\ e^{-\lambda M_\lambda} L(M_\lambda) & \geq e^{-\lambda(t^j - \tau^j)} L(t^j - \tau^j) \end{cases}$$

It must hold for any t^i and t^j, in particular, the ones which maximizes $e^{-\lambda(t^i - \tau^i)} L(t^i - \tau^i)$ and $e^{-\lambda(t^j - \tau^j)} L(t^j - \tau^j)$. The conditions become

$$\begin{cases} e^{-\lambda M_\lambda} L(M_\lambda) & \geq pL(M) - c \geq e^{-\lambda M_\lambda} L(M_\lambda) \\ e^{-\lambda M_\lambda} L(M_\lambda) & \geq e^{-\lambda M_\lambda} L(M_\lambda) \end{cases}$$
$$\Leftrightarrow e^{-\lambda M_\lambda} L(M_\lambda) = pL(M) - c$$

2) If $\tau^i < M_\lambda$ and $M_\lambda \leq t^i$, the conditions for this equilibrium to exist become

$$\begin{cases} e^{-\lambda M_\lambda} L(M_\lambda) & \geq pL(M) - c \geq e^{-\lambda(t^i - M_\lambda)} L(t^i - \tau^i) \\ e^{-\lambda M_\lambda} L(M_\lambda) & \geq e^{-\lambda(t^j - \tau^j)} L(t^j - \tau^j) \end{cases}$$
$$\Leftrightarrow e^{-\lambda M_\lambda} L(M_\lambda) \geq pL(M) - c \geq e^{-\lambda(t^i - M_\lambda)} L(t^i - \tau^i)$$

The maturation delay $t^i - \tau^i$ which maximizes $e^{-\lambda(t^i - M_\lambda)}L(t^i - \tau^i)$ is still M_λ. Consequently, the condition become

$$e^{-\lambda M_\lambda}L(M_\lambda) \geq pL(M) - c \geq e^{-\lambda\tau^i}L(M_\lambda)$$

A necessary condition for this inequality to hold is

$$e^{-\lambda M_\lambda}L(M_\lambda) \geq e^{-\lambda\tau^i}L(M_\lambda)$$
$$\Leftrightarrow \tau^i \geq M_\lambda$$

As $\tau^i < M_\lambda$, it is impossible and the condition never holds.

3) If $\tau^i < M_\lambda$ and $M_\lambda > t^i$, the conditions for this equilibrium to exist become

$$\begin{cases} e^{-\lambda M_\lambda}L(M_\lambda) & \geq pL(M) - c \geq L(t^i - \tau^i) \\ e^{-\lambda M_\lambda}L(M_\lambda) & \geq e^{-\lambda(t^j - \tau^j)}L(t^j - \tau^j) \end{cases}$$

The publication time t^i which satisfies $M_\lambda > t^i$ and maximizes $L(t^i - \tau^i)$ tends to M_λ. Hence, the conditions become

$$e^{-\lambda M_\lambda}L(M_\lambda) \geq pL(M) - c \geq L(M_\lambda)$$

This is impossible as $e^{-\lambda M_\lambda} < 1$.

Hence, such an asymmetric equilibrium does not exist.

Bibliography

Bernard, Olga (2020) "Citizen science vs. traditional science: what drives the choice of the scientist(s)?", working paper.

Bobtcheff, Catherine, Jérôme Bolte, and Thomas Mariotti (2017) "Researcher's dilemma," *The Review of Economic Studies*, 84 (3), 969–1014.

Bobtcheff, Catherine and Thomas Mariotti (2012) "Potential competition in preemption games," *Games and Economic Behavior*, 75 (1), 53–66.

Brunnermeier, Markus K and John Morgan (2010) "Clock games: Theory and experiments," *Games and Economic Behavior*, 68 (2), 532–550.

Fudenberg, Drew and Jean Tirole (1985) "Preemption and rent equalization in the adoption of new technology," *The Review of Economic Studies*, 52 (3), 383–401.

Hendricks, Kenneth (1992) "Reputations in the adoption of a new technology," *International Journal of Industrial Organization*, 10 (4), 663–677.

Hopenhayn, Hugo A and Francesco Squintani (2011) "Preemption games with private information," *The Review of Economic Studies*, 78 (2), 667–692.

Lambrecht, Bart and William Perraudin (2003) "Real options and preemption under incomplete information," *Journal of Economic Dynamics and Control*, 27 (4), 619–643.

Lukyanenko, Roman, Jeffrey Parsons, and Yolanda F Wiersma (2016) "Emerging problems of data quality in citizen science," *Conservation Biology*, 30 (3), 447–449.

Polymath, DHJ (2012) "A new proof of the density Hales-Jewett theorem," *Annals of Mathematics*, 1283–1327.

Weeds, Helen (2002) "Strategic delay in a real options model of R&D competition," *The Review of Economic Studies*, 69 (3), 729–747.