Research Design and Research Systems:

An Application of Agent-Based Modelling to Research Funding

A thesis submitted to the College of Graduate and Postdoctoral Studies
in partial fulfillment of the requirements for the degree of Master's in Public Policy (MPP)
in Johnson Shoyama Graduate School of Public Policy (JSGS),
University of Saskatchewan,
Saskatoon, Saskatchewan

By:
Ebrahim Hassanpour

PERMISSION TO USE

In presenting this thesis in partial fulfillment of the requirements for a Graduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis work, or in their absence, by the Director of the Johnson Shoyama Graduate School of Public Policy (JSGS) or the Dean of the College of Graduate and Postdoctoral Studies. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be permitted without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis/dissertation.

Ebrahim Hassanpour

Requests for permission to copy or to make other use of material in this thesis/dissertation in whole or part should be addressed to:

Director
Johnson Shoyama Graduate School of Public Policy (JSGS)
101 Diefenbaker Place
University of Saskatchewan
Saskatoon, Saskatchewan S7N 5B8
Canada

OR

Dean
College of Graduate and Postdoctoral Studies
University of Saskatchewan
107 Administration Place
Saskatoon, Saskatchewan S7N 5A2
Canada

ABSTRACT

Governments nurture their multi-disciplinary innovation systems by funding several public organizations to help universities and research institutes support research projects and associated infrastructure. To study the impact of research funding, a generic stylized model is developed using Agent-Based Modelling (ABM) to simulate the outcomes. To provide context, the analysis anchors the problem in the context of Genome Canada's research funding efforts. The process of academic research and the impact of grants on its speed and output (papers published) is simulated. To compare the outcomes for policy choices, two measures or indices are developed for the outcomes: efficiency is measured by number of papers per granted money and equity is measured by a Gini coefficient (for papers and money granted); the Matthew effect is also tested to check for effects on equity. Defining academic investigators as the main agent and having investigations and grants as subagents, along with assumptions for the procedures and parameters, an ABM is designed in which investigators conduct individual research using grant and non-grant funds. The simulation model is then tested and verified to be used for evaluation and comparison of policy scenarios. The results revealed that the instruments of allocated budget per competition, the gap between competitions, the sum granted for any proposal, and the size of the target group may be utilized to improve the efficiency and equity of the system. However, there is usually a trade-off between these two objectives and a loss in one of them is necessary to achieve a gain in the other. The tools can be combined in order to secure better results, but there are other factors that should be taken into account in making decisions. Although some lessons can be learned from such a simple model, making it applicable to policy making and to real-world issues, other factors such as investigator heterogeneity, collaborations, and grant administration complexities should be taken into account.

ACKNOWLEDGMENTS

For the idea and completion of this thesis I am indebted to some persons to be acknowledged below:

I would first like to acknowledge the generous financial support I received over the course of my studies from the VALGEN and Saskatchewan Innovation and Opportunity Graduate Scholarship. I want to express my sincerest gratitude to my supervisor, Peter W.B. Phillips, for his kind mentorship and support since the beginning of the idea. The idea was his and I am very grateful to him for taking me into a completely new field, the simulation and coding. I have learned a lot from him both for my thesis and my life as well. Without his encouragement, I would not be able to finish.

I would also like to thank Nathaniel Osgood for his help in simulation and coding. Although I was not lucky to work with him till the end, I learned a lot both in his simulation bootcamp and class. The idea of switching to AnyLogic was his and it was a great step. I further thank my committee members Kathy McNutt and Justin Longo and the external examiner, Samantha Evans for their time and valuable comments. I must acknowledge a couple of friends who helped me with coding including Rahim Oraji, Farhad Maleki, Amir Sadeghian, Dylan Knowles, Mohammad Shabani, and Aydin Teyhouee.

Finally, and most especially, I am grateful for my family (my wife and children) for their support throughout the whole course. There are others to be thanked including the faculty members of JSGS and other friends and family members whose names would make the list too long.

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES

Chapter 1:

INTRODUCTION

As a key element in seeking economic prosperity and superiority, science and technology (S&T) policy addresses the public-sector measures designed to create, fund, support, and mobilize scientific and technological resources (Arvanitis, 2003). A typical S&T policy includes public activities such as investing directly in R&D, being involved directly and indirectly in business R&D and other innovation promoting measures. As with other industrialized countries, the Canadian government seeks to improve its science, technology and innovation system. Its current strategy seeks to improve Canada's competitiveness through investments and activities in three key areas: the role of the private sector in innovation, research excellence and strategic R&D, and knowledge-based workers. The current priority areas in technology are environmental science; natural resources and energy; health and life sciences; and Information and Communications Technology (ICT) (OECD, 2012). Canada, relative to other Organization for Economic Cooperation and Development (OECD) member countries, relies much more on research undertaken in the higher education sector. Given the relative importance of scholarly research for economic development, there is value in investigating the design and dynamics of research funding and research outcomes in that space.

Research funding is one the tools governments use in their S&T policy. There are multiple methods to fund research—the most important is government grants. Public grants are given to researchers to spend on investigations which are hoped to result in outcomes that will improve the welfare of society. Deciding on who gets the grant money and on the size and other attributes of the grant are among the policy tools in research funding. These decisions certainly affect the outcomes of the policy and knowledge about the impacts are necessary for better policy making.

A generic stylized simulation model is developed here which studies the basic elements of academic investigation and the impact of some policy parameters on the outcomes. To provide context for the model, Genome Canada is explored to illustrate the applicability and working of the model. Below, the Canadian innovation policy is reviewed with an emphasis on Genome Canada followed by an articulation of the problem statement.

1.1 Canadian Innovation Policy and Genome Canada

The Canadian government nurtures its multi-disciplinary innovation system, by substantially funding several public organizations to help universities and research centres to support research projects and associated infrastructure, to develop talent, and to create collaborative research and development networks (STIC, 2015). The most important of these organizations that receive federal investments in S&T policy are the National Research Council Canada (NRC), the Canada Foundation for Innovation (CFI), the Natural Sciences and Engineering Research Council (NSERC), the Canadian Institutes of Health Research (CIHR), the Social Sciences and Humanities Research Council (SSHRC), the Networks of Centres of Excellence of Canada (NCE), the Canada Research Chairs Program (CRCP), the Canada Excellence Research Chairs (CERC) and the Canada First Research Excellence Fund (CFREF). Along with an array of public institutes and universities whose research they support, and along with provincial research organizations, the above-named organizations form the public part of the Canadian innovation system.

A relatively new addition to the above list was Genome Canada (GC), with a national office and six regional centres. The growing importance of research involving genomics[1] led the federal government to create the five-year Canadian Genome Analysis and Technology (CGAT) program in the 1990s, which was later replaced by Genome Canada in 2000 (Genome Canada, 2010). As a non-profit organization, GC serves as a catalyst for developing and applying genomics and genomic-based technologies to create economic and social benefits (Genome Canada, 2016). To this end, GC connects ideas and people across public and private sectors to find new uses for genomics, invests in large-scale science and technology to fuel innovation and translates discoveries into solutions (Genome Canada, 2016).

The founding of GC was part of a broader governmental objective often referred to as "Canada's national innovation strategy," which also led to the creation of the CIHR and the CFI among other research foundations (Hinterberger, 2010). Created and incorporated under the Canada Corporations Act, GC's mandate is to develop and implement a national strategy in genomics research for the benefit of all Canadians, by investing in large-scale genomics research initiatives in sectors of strategic and economic importance to Canada (i.e., health, agriculture, environment, forestry, fisheries, energy and

[1] Genomics is an area within genetics that concerns the sequencing and analysis of an organism's genome. The genome includes the entire DNA content that is present within one cell of an organism. (ISED, 2015).

mining), by aiming to strengthen genomics research and technical capacity in Canada, and by fostering multi-sectorial partnerships nationally and globally (ISED[1], 2015).

GC also works to ensure that genomics research considers underlying ethical, environmental, economic, legal or social aspects (GE^3LS) and that the research provides Canadian scientists with advanced technologies and expertise for funded projects by supporting the operations of five Science and Technology Innovation Centres[2]——STICs (ISED, 2015). Genome Canada delivers its mandate through six Genome Centres in British Columbia, in Alberta, on the Prairies, in Ontario, in Quebec and in the Atlantic region. These centres administer the funds to research projects and are responsible for identifying regional strengths and opportunities, monitoring compliance and performance, and helping secure co-funding from partners.[3]

According to ISED, the main sponsor of the program, Genome Canada has received almost $1.5 billion from it and has raised over $2.1 billion through co-funding commitments (Genome Canada, 2017). The co-funding partners include provincial governments and agencies, international non-governmental organizations and research institutes, industry, universities, and research hospitals. Structurally, ISED contributes funding directly to Genome Canada, which then launches national competitions and a merit review process to select the research projects it will support. Once the selection is completed, Genome Canada in turn funds the Genome Centres and STICs, who then transfer the appropriate funding to selected research projects.

While ISED monitors Genome Canada's program, a Board of Directors with representatives from the academy and industry along with a Chief Executive Officer (CEO) manage it. Previously, GC's co-funding commitment was based on a 1:1 ratio (between funds of ISED and those from other sources). Now, for every dollar provided by ISED, GC must raise two dollars from other sources, in a 1:2 ratio (GC, 2016). The target population served by the program is the genomics research community located

---

[1] Industry Canada has changed its name to Innovation, Science and Economic Development Canada (ISED) since July, 2017. Therefore, Industry Canada has been replaced by ISED in the thesis.

[2] These centres have been changed into ten technology platforms (see, https://www.genomecanada.ca/en/about-us/genomics-technology-platforms)

[3] It should be noted that there exists another organization with a similar mission, called the Genomics R&D Initiative (GRDI). The GRDI was established in 1999 to build and maintain capacity inside government departments (initially six but now eight federal government organizations) and to carry out and support genomics research according to their respective legislative, regulatory and policy mandates (Genomics R&D Initiative, 2007 and 2014).

in Canadian universities, research hospitals, and non-profit research institutions. The funds are granted to the target population through a competitive process in a manner to achieve the most.

The principal focus of GC's research program is large-scale research carried out by teams of researchers bidding on multi-year, interdisciplinary research contracts (Doern et al., 2016)., GC has, then, conducted periodic competitions to fund genomics-related R&D projects. Since the start of the program in 2000, GC has administered seven major competitive research competitions and almost 20 smaller-scale, more-focused research initiatives. While the first three competitive calls were open to any applicant, the last four were targeted on specific domains. Table 1.1 presents information on the competitions held.

Table 1.1: General information on the funding competitions held by Genome Canada

Competition #	**Year**	**Area of Focus**	**GC Funding** (millions of dollars)	**Number of Projects Funded**
I	2000-1	All	81	17
II	2001-2	All	146	34
III	2004-5	All	205	33
ABC	2008-9	Bioproducts & Crops	53	12
LSARP I	2010	All	29.9	7
LSARP II	2010	Forestry & Environment	30	9
LSARP III	2012	Personalized Health	45.1	17
LSARP IV	2014	Feeding the Future	30.8	11
LSARP V	2015	Natural Resources & the Environment	26	NA
Other	2003-15	19 Narrow calls	>300	90*

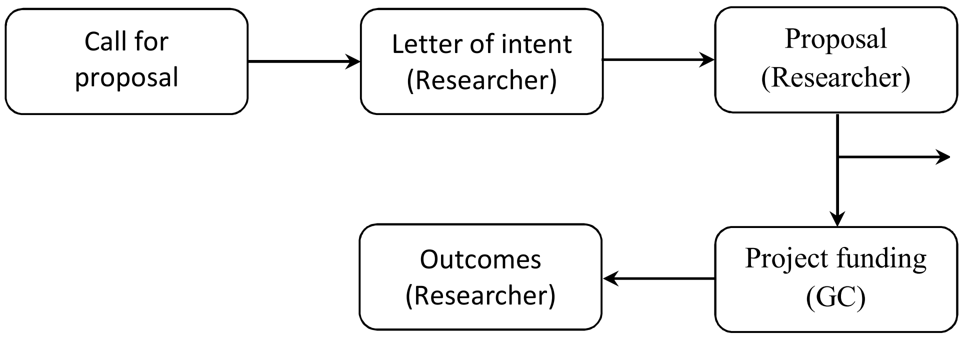
* Estimate

Adapted from: Doern et al. (2016), Sharma (2013) and GC (2016)

Since the funds could not be disbursed completely in the first round in 2001, a second round of competitions was held in the same year. A total of 51 projects were funded in 2001 with a budget of over $290M. (Within the year 2001, GC received $160M first as a grant from the federal government and, shortly afterwards, received another $140M, of federal government funds. For this reason, Competitions I and II were held in the same year.) A few years later in 2005, 33 projects were funded with a budget of $346M, before funds dropped to $112M (awarding only 12 projects) in 2008-9, which confined the funding to specific agricultural areas (the competition was called Applied Genomics in Bioproducts and Crops, ABC). Since 2010, Genome Canada has turned to large-scale applied research projects (LSARP), focusing on specific domains at every round of funding. GC's share of the funds is presented in Table 1.1 and, according to GC (2016), the Federal government, through Innovation, Science and Economic Development Canada, has committed $1.2 billion in funding to Genome Canada since 2000-01, of which approximately $1.1B has been allocated to projects and initiatives and $94M has been spent on operations. The total funding by GC and co-funding agencies has therefore been about $2.7B.

GC competitions usually start with an announcement (call for proposals), to which scholars apply, in letters of intent (LOI). When the initial approval is granted by GC, the leading or principal investigators start forming teams and writing proposals to secure GC grants. The proposals undergo an evaluation process and upon approval, GC determines the size and method of funding. The funds are allocated and the investigations are carried out at respective universities or other research centres. When research projects are completed, scholars report their outcomes such as whether or not papers were presented and/or published, whether inventions or innovations were patented, and whether the training of human capital occurred, in the form of students or researchers. Figure 1 gives the basic map of the process.

Figure 1.1: Schematics of project funding decision process by Genome Canada



To improve the research and application of genomics in Canada, GC must make choices about some parameters in the granting of funds. For instance, the size of the fund for each round, the size and diversity of the project's team members, the target population, and the maximum grant for any single project are among GC's policy variables. In order to make an informed decision about these variables, some knowledge about their impacts on the outcome is necessary.

1.2 Research Objective

As mentioned above, the main objective of this thesis is to develop a stylized simulation model for academic research funding. The method used is Agent-Based Modelling (ABM), which simulates the process of academic research and the impact of grants on the speed and output of that research. Watts and Gilbert (2014) argue that "simulation models of science are an excellent basis for studying innovation processes" (p. 135), and that innovation is best approached as a 'social and collaborative' process and that data trails are left that can be used in tracing the process, mechanisms and patterns. Simulation can be used for finding the mechanisms that give rise to such patterns and thereby help policy-makers to improve the design and function of the system.

Here, the modelling and simulation are performed step-wise from the simplest case to a case close to reality. Some complex stages like networking and team-formation are left for further research. Although outcomes of research are beyond papers and patents, and training is also important, only the papers will be considered as the output. To compare the outcomes for policy choices, indices or measures of the outcomes are needed. In this thesis, two measures are developed to assess outcomes: the number of papers per granted unit of money invested, which measures efficiency; and equity as measured by a Gini coefficient (for papers and money granted). The policy options will be compared by the outcome measures making it possible to draw policy implications generally and, by implication, for organizations like Genome Canada.

The model can be used in any research funding context with some adjustments for the specific situation. Although investigations are assumed to be carried out individually without collaborations, the basic results should be valid. While this is a stylized model neglecting many features of the real policies. It can be adapted to real-world policies by use of more precise institutional factors and policy information.

Chapter 2:

LITERATURE REVIEW

2.1 ABM in Social Sciences

Agent-Based Modelling (ABM) which is also called Individual-Based Modelling (IBM) in fields such as ecology (see e.g. Grimm and Railsback, 2005; and Railsback and Grimm, 2012) and Multi-Agent Modelling by some authors (see e.g. Gilbert and Troitzsch, 2005), was originally developed in the 1950s (Urban et al., 2011). Since an extensive use of computer software is made to show the dynamics of the agent behaviour, ABM is called a computational method and, as Nigel Gilbert describes, "agent-based modelling is a *computational* method that enables a researcher to create, analyze and experiment with models composed of agents that interact within an environment" (Gilbert, 2008, p. 2). ABM is a modelling and simulation method in which the interaction of the agents (human or non-human) is taken into account and this feature, along with its bottom-up construction, distinguishes it from other simulation methods like System Dynamics (SD) and Discrete Event Modelling (DEM). When applied to the social sciences, ABM is also referred to as a form of computational social science (for the realm and methods of computational social science, see Bankes et al, 2002; Conte et al, 2012).

The first application of ABM to the social sciences is considered to be Thomas Schelling's segregation model. Schelling (1971) used the simulation approach to illuminate racial segregation in U.S. cities. Schelling's main concern was the emergence of results from individual behaviour, where the attitude of agents (households) towards their neighbours' race would result in racial separation of the neighbourhoods. A decade later, Robert Axelrod (Axelrod, 1981) applied the technique to understand the cooperation of players in a prisoner's dilemma game in the context of Darwin's evolution theory. The availability of simulation computer software such as NetLogo, paved the way for extensive application of ABM in the 1990s in various disciplines like sociology, political science, economics, and business; some cases have been reviewed in Gilbert (2008). According to a review of the literature on applying ABM to the social sciences, out of the 279 journal articles published during 1998-2008, applications to economics ranked first with 29 percent of the papers followed by those to social science (24%), biology (14%), military studies (13%), and public policy (8%) (Heath et al., 2009). The establishment of the multidisciplinary *Journal of Artificial Societies and Social Simulation* (JASSS) in 1998 was another major step forward.

Toward the end of 20th century, some scholars became interested in applying ABM to innovation and technological change. The first application of ABM in this context (without mentioning the term, but just calling it a "simulation model") was Gilbert (1997), which simulated the structure of academic science and concluded that "it is possible to generate many of the quantitative features of the present structure of science …." Other earlier studies were carried out in the realm of *evolutionary economics* and used some kind of simulation of the innovation process (see e.g., Ballot and Taymaz, 1997, 1999; Cantner and Pyka, 1998). Starting in early 2000s, the EU project of Simulating Self-Organizing Innovation Networks (SEIN) and its SKIN (Simulating Knowledge Dynamics in Innovation Networks) model used ABM simulation extensively. The results of those studies have been being published since then: some of the papers will be reported in the next section. Having carried out an almost complete review of the application of ABM (called Agent-based Computational Economics or ACE, in that context) Dawid states that "despite the apparent merit of the agent-based simulation approach for the analysis of a wide range of issues in the economics of innovation and technological change, the amount of relevant ACE-based work in this area is not huge" (Dawid, 2006, p. 1242).

2.2 ABM in Innovation Context

Modern innovation is a collective process taking place in universities and research centres (public or private). The network of researchers is of a great importance to this process. Accordingly, the study of innovation processes can be carried out from different perspectives, including process simulation, network analysis and other socio-economic or motivational viewpoints. Since the current study emphasizes simulation and networking aspects, a review of the relevant literature is presented below.

The first attempt to simulate formal scientific activity is Gilbert (1997). After defining and classifying simulation models, he simulates Lotka's Law, which states that for scientists publishing in journals, the number of authors is inversely proportional to the square of the number of papers published by them; a distribution called Zipf applies for this phenomenon. Considering science as an evolutionary process, Gilbert uses *papers* and *authors* to characterize the institution of science in which each paper brings a new quantum of knowledge. In order to represent a quantum of knowledge, he uses a sequence of bits and calls those sequences *kenes* –in analogy with *genes*. A kene is the knowledge contained in a paper and that kene can represent the relevant paper. Gilbert writes: "Papers are generated from other papers, sharing a kene with their generator, but modified according to the kenes of the papers which it cites" (paragraph 8.2). Assuming two coordinates for the kenes and placing time as the third dimension, the model is run for 1000 time steps and is noticed that papers cluster in some spots. Lotka's Law regarding the distribution of citations among authors is reproduced and many of the features of the academic science structure are generated, by using their simulation.

Interest in the simulation of innovation processes dates to the late 1990s when Cantner and Pyka (1998) develop a model of technological progress at the firm level within the context of a knowledge-based (contrary to a Neoclassical, resource-based) approach. Using this approach, they simulate the firms' technological trajectory. A few years later (Pyka, 2002), motivated by a lack of detailed study on the theory of emergence and diffusion of innovation and its networks, Pyka uses the knowledge-based approach of evolutionary economics in an attempt to develop an evolutionary theory of innovation networks. He argues that the important features of inter-firm learning and synergies in innovation networks cannot be captured by a traditional, incentive-based approach.

Teitelbaum and Dowlatabadi (2000) use the ABM simulation approach to study the innovative behaviour of heterogeneous firms and their interactions. In their model, firms allocate resources to R&D activities leading to radical or incremental innovations and new products. Their model consists of products (represented as binary vectors with 12-bit length like a genetic code, for product attributes and the skills needed to produce them), firms (profit-seeking and competing-over-market-share), consumers, and an environment (a lattice for locating agents). Three types of simulations are performed: ideal strategy distribution where firms have complete information about each other's activities; effect of spillovers in which they do not have access to information of others' innovative attempts; and adaptive firms in which learning and mimicking successful firms is allowed. The study shows that in every case, heterogeneity in firms results in better performance of the industry, implying that a synergy exists among firms of different innovation strategy.

Building upon previous studies and emphasizing the importance of networks in innovation, Gilbert et al. (2001) develop a simulation in which the networks "evolve from the dynamic and contingent linkage of heterogeneous units each possessing different bundles of knowledge and skill" (paragraph 1.2). The European Self-Organizing Innovation Networks (SEIN) project[1] is used as the case study. Ahrweiler

[1] The SEIN (Self-Organizing Innovation Networks) project is a combination of agent-based simulation of five case studies in technological innovation in different European countries, started in late 1990s. Its model is called SKIN (Simulating Knowledge dynamics in Innovation Networks) and it is an ABM model used by the European Commission for scenario modelling of current and future innovation policy strategies (CRESS). The model is "a multi-agent simulation of firms that try to optimize their innovation performance in order to respond to the requirements of a constantly changing environment" (Gilbert et al, 2007, p. 100).

et al. (2004) use an extended version of the SKIN model to study inter- and intra-firm knowledge dynamics. It uses Gilbert's kene idea but here to represent the knowledge base of a firm composed of units of knowledge. A triple of *capability* (C), *ability* (A), and *expertise* (E) represents a unit of knowledge and a series of these triples form a kene. These firms' initial capital is used to produce outputs and to increase the knowledge base. A subset of a firm's kene triples, which becomes the focus of its innovation, is called an *innovation hypothesis*. The hypothesis can be developed into a product using C and A and whose quality depends on a combination of A and E. In the same manner, inputs are produced and supplied by other firms and traded on the market. The production in any round increases the firm's expertise (by one unit) in its innovation hypothesis (learning by doing) and the E level in other triples decreases by one. These firms can better their performance in an individual or cooperative manner and incrementally or radically; they may live in a closed, only-firms world or in an open environment with external suppliers and buyers. The formation of partnerships or collaborations starts when the firms search for potential partners by studying the capabilities of others that are disclosed in their advertisements. This way, the partners copy their different triples and, when the partnerships are profitable, a network can be formed. Start-ups are encouraged by the profitability of their businesses and are represented by adding new agents. The model is run in NetLogo using 100 agents.

Gilbert et al. (2007) use the SKIN model (the same version used in Ahrweiler et al, 2004) to study organizational learning and its different strategies. Emphasizing that the capacity of firms to learn determines their competitiveness, different forms of learning by firms (by doing, by feedback, adaptation/incremental, and innovative/radical approaches) are modelled and their impact on the industry is considered. They start the simulation simply "by doing" and by receiving feedback, then add adaptive learning (through incremental research), then add innovative learning, and finally attain external knowledge through partnering and networking. The simulation reveals that a combination of internal research and partnership produces the best result (more innovation and fastest growth in the populations of the firms). Gilbert et al. (2007) conclude that in highly dynamic modern knowledge-based businesses, learning by doing and learning by feedback, are not enough for firms to survive and should be combined with other forms of learning.

After writing a critical overview of the theory of industrial organization in dealing with R&D collaborations, Pyka et al. (2007) study innovation networks and their development using ABM simulations. The same previous SKIN model is used, but the simulation results are compared with empirical findings from the UK and German bio-pharmaceutical industries, in order to calibrate the model to real world observations. In the standard scenario, the situation in knowledge-intensive industries is simulated, where innovation networks are prevalent. The model starts with 500 firms of which 50 of them are large. The number of firms (mostly small ones) and the number of networks fluctuates during the simulation time with an increasing trend. Looking more closely at the networks reveals that most of them are small with a few actors, but a couple of large networks exist which are very important for knowledge flow. The next step is to change some parameters (the initial size distribution of firms, cooperation strategy that is conservative among similar-knowledge firms or progressive among diverse firms; and an attractiveness threshold which measures how quickly actors decide to cooperate) to build scenarios that are comparable to empirical findings. The conservative strategy scenarios give rise to scale-free networks (or power log distribution of networks) which are a feature of real-world innovation networks in knowledge-intensive industries.

Four years later, the SKIN model is used to study the academia-industry links. According to Ahrweiler et al. (2011a), although there is a lot of theoretical evidence supporting the positive impact of universities' presence on industry innovation, there are opposing empirical findings about links between universities and industry. The study applies the ABM to investigate these links and their impact on innovation generation and spread in industry networks. The SKIN model is extended by adding non-profit, homogeneous university agents, which enjoy more knowledge than firms do and are modelled with longer kenes; these agents were assumed to link only with other firms without knowledge flowing to the linked ones. To see the impact of the presence of the universities in the model, two scenarios, one with and one without university links, are simulated and compared using a statistical *t*-test. Based on the previous studies and arguments, four hypotheses are developed and tested regarding the impact of universities on the performance of the firms and of the whole system. Results show that university cooperation increases the knowledge and competence levels of the industry, enhances the variety of knowledge among the firms, and improves the quantity and velocity of innovation diffusion. Also, firms interacting with universities are found to be more attractive to other firms, when new partnerships are sought.

There is a long-standing debate on the relative importance of strategies of individual actors (agency-oriented patterns) and institutional frameworks of innovation systems (structure-oriented patterns) for the performance of those innovation systems. Ahrweiler et al. (2011b) use the same SKIN model to investigate the relative significance of these internal and external factors for innovative success (number of innovations) and for the size (number of firms) of the industry. For the agency-oriented scenario, Ahrweiler et al. test whether the strategic collaborative decisions of actors are responsible for shaping the sector; the researchers also use the permanently changing distribution and (re)combination of types of capital to represent structural conditions. Using regression analysis of the data generated from simulations, Ahrweiler et al. show that not only the capital distribution but also the partner choice mechanisms matter for the innovation performance. Hence, neither the agency- nor structure-oriented scenarios are supreme and both are needed if an efficient system of innovation is sought.

To see how ABM methodology can contribute, through a learning-by-modelling process, to increase our knowledge of complex phenomena, Triulzi and Pyka (2011) use a model of university-industry relationships (UIRs) in the biotech and pharmaceutical sectors. They argue that there is no consistent and complete evidence regarding the long-term impacts of UIRs on the innovativeness of the research system. Multi-agent modelling is used to reproduce the dynamics and productivity of such relations. Both university and firm (either diversified or specialized biotech) actors are involved to perform research, along with two types of funding agents (national research agency and venture capitals). The knowledge kene concept is used, but here with a fourth element of research direction to differentiate universities from firms (basic vs. applied research). Research projects are simulated to be carried out individually or jointly and upon success, lead to patents which directly or by licensing, create new drugs and increased revenues. These results show that UIRs increase incentives for universities to engage in applied research and cause a significant increase in the innovative potential of biotech firms. In the biotech and pharmaceuticals sector, public and industry research funding complements but does not substitute for each other. The simulation is considered through a double-loop, learning-by-modelling process, which generally starts from theory and is refined through a verification and validation cycle, in which the theory itself is confirmed and improved. Finally, Triulzi and Pyka argue that "the ABM methodology can substantially contribute to a better understanding of complex socioeconomic interactions and thus support the development of theories that are suited to dealing with this complexity without ignoring it" (Triulzi and Pyka, 2011, p. 498).

2.3 Conclusion

Reviewing ABM application to the field of innovation policy and practice shows that the literature, which began to be published in the late 1990s, has mainly concentrated on European SEIN projects which are focused on specific industries. In the Canadian innovation space, although there are studies on innovation networks (see e.g. Sharma, 2012; Boland et al, 2012; and Ryan et al, 2014), ABM simulation is still absent. The current study represents an initial attempt to apply the popular, agent-based simulation to a Canadian innovation policy, i.e. Genome Canada program. Once this simulation has been built, its basic model can later be improved and adapted to simulate different scenarios and policy contexts.

Chapter 3:

METHODOLOGY

To build a model that applies to the real world, a researcher must first describe the situation that the model will account for. The following section describes the process of academic research and the role that grants play in that process. Following this description, we will examine a graphical model that lays the groundwork for agent-based modelling, which takes place through computer visual modelling. The definition of the terms and agents necessary for the modelling and simulation are presented in the same sections. The simulation software is AnyLogic 7.

We will follow a protocol for ABM here. Proposed in 2006 by Grimm and 27 modelers from various countries and universities, this protocol sets some rules for describing simulation models, since they are structurally more complex than analytical ones. In other words, the protocol suggests a standard for publishing ABM simulation to make it understandable and replicable. The protocol is called ODD (Overview, Design, and Details) and consists of elements relevant to these three items. The basic protocol and its update (Grimm et al., 2010) are followed in this study.

3.1 Academic Investigation and Grants

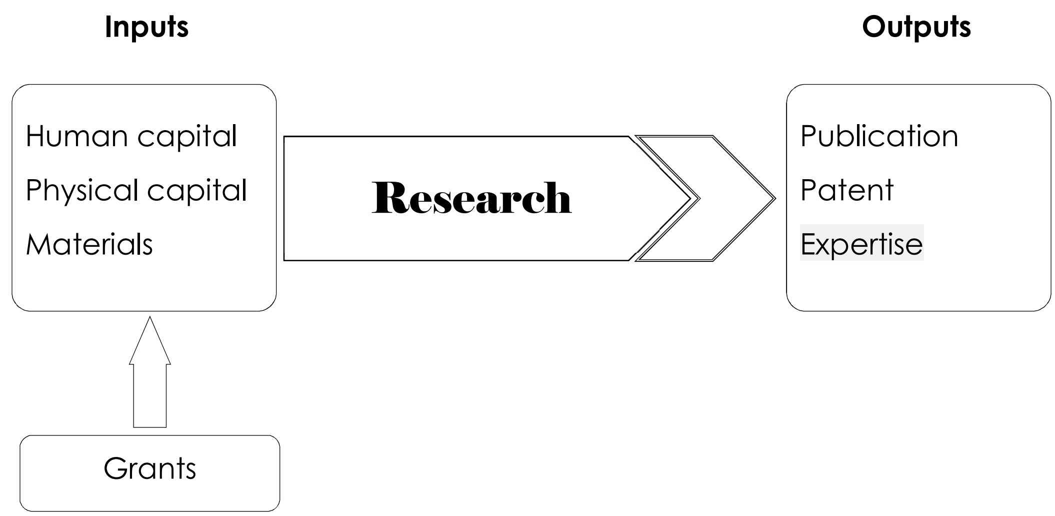
Most current academic research is performed in university departments, where professors do the research either individually or in collaboration with colleagues and students. Although to some extent faculty members may fund their own line of investigation, the financing of academic research is mainly carried out by external agents. Formal research funding takes place either through research grants, research contracts, or a combination of the two (Hakim, 2000). The main difference between grants and contracts lies in who has the authority over, responsibility for, and the control of, the project funds. Research contracts, as the name implies, take the form of contracts between two parties where the investigator (contractor) receives the funding in order to find answers for the questions posed by the contractee. The conduct of the research is a joint responsibility and the contractee has some authority over the research process and outcomes. By contrast, in research grants, the researcher has sole responsibility for the design of the study, any modifications made to it, and the implementation of the project (Hakim, 2000). This difference between contracts and grants, in the authority and responsibility of their agents, causes industry usually to outsource R&D work to university faculties through contracts, while government agencies outsource R&D projects mostly through grants (Bozeman and Gaughan, 2007). This difference of government involvement explains why grants (especially public ones) are more popular in academia.

There are a number of other classifications for grants (see Auranen and Nieminen, 2010; Huffman and Just, 1994). In the case of matching grants, the investigator would be responsible for finding another (local) agency to match the contribution of the main grantor. Formula granting is used in cases where a limited amount of money is to be divided among investigators (such as faculty members), using some pre-established rules or a specific formula. By contrast, competitive grants involve a competition process where only some of the participants win. The focus of this study will be on competitive grants rather than other forms of grants and contracts. However, the modelling of the processes and outcomes would be similar across these classifications.

Academic research is partly conducted as a learning process for graduate students, which does not always depend on external funding. In such a case, laboratory resources of the departments along with the intellectual capital of professors and students may result in publications (usually journal papers) or in innovations or inventions (whether patented or not). The learning and expertise acquired by the investigators is a side product. In such a process, external funding acts as a catalyst to speed up the process and also to implement some large research projects which in the absence of funding would not have been accomplished. In short, academic research is a production process that begins with human and financial input and ends with the output of papers and patents. The practice of learning-by-doing promotes the quality of the human resources involved.

Figure 3.1 provides a chart that explains the research process. Researchers and their associates compose the human capital which employs physical capital (equipment, machinery, tools, field/office) and materials (anything consumed in the process such as chemicals, paper, energy, etc.). The only resources used up in the process are materials and only the services of human or physical capital are used. For this reason, some experts instead prefer to use the term "human and capital services" as the inputs. The research process itself involves various stages (literature review, method development/prototype design, experiment and data collection, analysis, and report writing/documentation) that, depending on the nature of the R&D, may need a time span of weeks to years, in order to be accomplished. The findings of the research are reflected in publications (paper, book, or report) or realized in inventions; as mentioned before, the learning achieved during the process also adds to human capital. Grants, and funding in general, are used to provide inputs by compensating for human resources, for buying or renting tools, equipment, and offices, and finally, for providing necessary materials.

Figure 3.1: Schematic presentation of academic research



3.2 ODD Presentation of the Model

Although the ABM model can be presented in the order of its development or in any other logical order, some simulation experts prefer it to be presented specifically, according to ODD protocol. Therefore, that is followed in this study.

3.2.1 Overview

a) Purpose

The simulation will be used to study the effect of different structures of research design on the outcome of its research grants.

b) Agents and state variables

The principal agent in the model is the investigator (researcher). These agents are busy with academic investigation related to genomics and proteomics and related social science fields. The investigations happen individually and no collaboration is taken into account. There are two dependent sub-agents for the investigator: Investigation and Grant. The investigator is always busy with at least one line of investigation which results in the output of a paper. The investigations and the resulting papers are saved for any investigator. The investigators take part in every granting competition, with some of them winning and some receiving rejections. The grants won make feasible more and new investigations that result in more output for the winners. The paper output of the investigators is taken into consideration by the granting agency, when it assesses grant proposals, in the next rounds. Beside the score the granting agency gives to a paper, a random score is considered to account for output quality and pertinence. The granting agency has its own timeline of application and life which will be explained in the processes. Table 3.1 lists the agents along with their state variables.

Table 3.1: Characteristics of the model agents and their attributes

Agent	Sub-agent	State variable	values
Investigator		Region	6 regions (BC, AB, PR, QC, ON, AT)
		Paper score	1-4 per paper
		Random score	≤ paper score
		Total score	Sum of paper and random scores
		Annual fund without grant	\$2000/ month
	Investigation	Duration	24 months
		Paper	1-5
		Cost of an investigation	\$2000/ month
		Publication duration	3-12 months
		Paper quality distribution	1-4
	Grant	Grant writing time	6 months
		Grant assessment time	1 month
		Grant size	\$120-500 thousand
		Budget for every competition	\$50 million
		Competition gap	36 months
Environment parameters		Number of investigators	600
		Simulation horizon	26 years

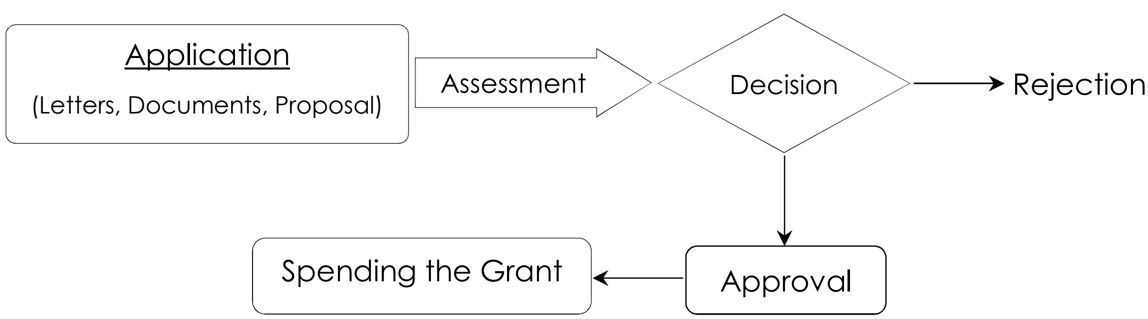
Source: Research findings

c) Processes

Investigators use their resources to perform R&D and to produce outputs (papers). Every faculty member follows a few lines of research during their academic lives. They work with students who help them in the process and who, at the same time, receive some training. Investigators use some form of physical capital in the form of offices, labs, tools, equipment, etc., and there are some materials consumed in the process. All of these capital services and materials require money in order to be purchased and that is where grants are needed and used.

Various sources and kinds of grants are available for researchers but they are obtained through a process. Finding an idea and gathering some information about it is the first step. Then comes the writing of an application, which is normally submitted in the form of a proposal. The proposal is reviewed by the funding agency and, if accepted, the grant money is given to the applicant. Each of these steps requires some time and in the end, some research takes place that results in a particular output. Grants make it possible to start new investigations. The grant process is depicted in Figure 3.2.

Figure 3.2: Schematic presentation of grant process



3.2.2 Design Concepts

Agents are randomly spread across six non-identical regions all across Canada. The agents compete for grants provided by the granting agency. At the first round, the winning is quite random, giving an opportunity to the winners to publish more papers. The paper output (and score) is used in subsequent rounds as a criterion to assess grants, which provides a better chance for the winners in future. As a result, in spite of the homogeneity of agents in the beginning, some inequality and distinction emerge in the process. The agents are assumed to be researching by themselves and there is no interaction among them. Papers are assumed to be of different qualities, giving rise to some heterogeneity among those having the same number of papers. Also a random score is considered to account for the pertinence of papers for competition.

3.2.3 Details

a) Initialization

The simulation run starts with a population size of 600 (the approximate number of participants in the GC Applied Bio-products and Crops competition). The agents are spread across six regions, according to the distribution of university teaching staff in Canada. The parameter values are the same as those presented in Table 3.1, with some distributions being assumed for the range value. The number of papers per investigation follows a uniform distribution (from 1-5); the grant size is distributed triangularly with a min, mode, max of 120, 240, and 500 thousand dollars, respectively; the paper publication duration also follows a triangular distribution of 3, 6, and 12. The granting starts at the end of year 2 and continues every 3 years.

b) Input Data

The model uses some input data but is largely based on assumptions. For the distribution of the investigators across regions, the real data on the spread of full-time teaching staff in universities was used to get a custom distribution for the agent location. Also, the agent population is based on data from the last competition held by GC. Finally, for the scoring of papers, the spread of journals proposed by an Australian institute (Australian Research Council's ERA initiative) is used.

c) Submodels

As mentioned above, the principal agent in the model is the investigator agent who lives in an environment called Main. Inside any investigator exist two sub-agents: investigation and grant. The investigation is created by a process by the investigator when they can afford a complete one. Upon creation, the investigation passes through a *statechart* which starts with the *ongoing* state lasting for *investigation duration* after which it becomes *done*. After this *done* state, it takes some time picked from *publication duration* to reach an end which is some papers (defined by a uniform distribution of 1-5).

The papers are scored for quality according to a custom distribution called *paper quality distribution*. The sum of the scores of the papers is saved in *paper scores* for every investigator. This is one component of the score used to assess grant applications. There is also a second component called *random score*. This component is computed by multiplying a random number (0-1) by the highest paper score among investigators, leading to a number that is less than or equal to the paper score component. The sum of two components is called the *total scor*e, which is saved for every investigator and called upon grant application assessment.

Another *statechart* models the states and transitions for the grant sub-agent. Grant agents are created when each competition begin (with calls for proposal) which after *grant writing time* leads to the *application* state. It takes *grant assessment time* to transfer from *application* to a decision (choice), where the application is approved or rejected. The *approval* branch is decided by specific rules (random choice in the first round but score-based selection in subsequent rounds) put into a dynamic event called *grant assessor*. On approval, the grant enters *awarded and ongoing state* which lasts for *investigation duration*. The approved grants add the *grant money* they receive to the investigator's funds, which is then used to start new investigations.

The variable values are stored and updated in corresponding variables, dynamic parameters, or charts. Some events are used to perform calculations needed to get to outcome measures (described below). These measures along with the related graphs and charts are updated automatically until the simulation ends. Most of the results are directly stored and presented in AnyLogic itself (the presentation window). But a few of the results are entered into Excel to be further analyzed.

3.3 Outcome Measures

As a policy, every granting agency may consider two impacts of their efforts. The first aspect is efficiency associated with the cost-effectiveness of the system, which is here defined as how many papers are produced from the grants awarded. The second aspect is the equity dealing with a distribution of grant money and papers among the investigators. The distribution is important because the public does not want resources dispersed to only a few people, which may also have implications for efficiency. The outcome of the simulation is summarized by some measures that address these two aspects and that are used in model verification and scenario comparisons.

The efficiency of the system (or the productivity of the grants) is measured by papers per grant money; i.e.

$$PMD = \frac{number\ of\ papers}{grant\ money\ (millions\ of\ dollars)}$$

where PMD stands for papers per million dollars. This measure shows how efficient some policy or instrument combination is or how much better the system produces papers in a certain setting.

Equity aspects are caught by two kinds of indices borrowed from social sciences. The Gini index was developed originally in economics, where it was used to measure the distribution of income and wealth. The "Matthew effect" was theorized in sociology to explain a situation where some academic figures achieved unfair advantage as a result of some initial fame or chance resulting in a growing gap among faculty members (or scientists) in the same discipline or field.

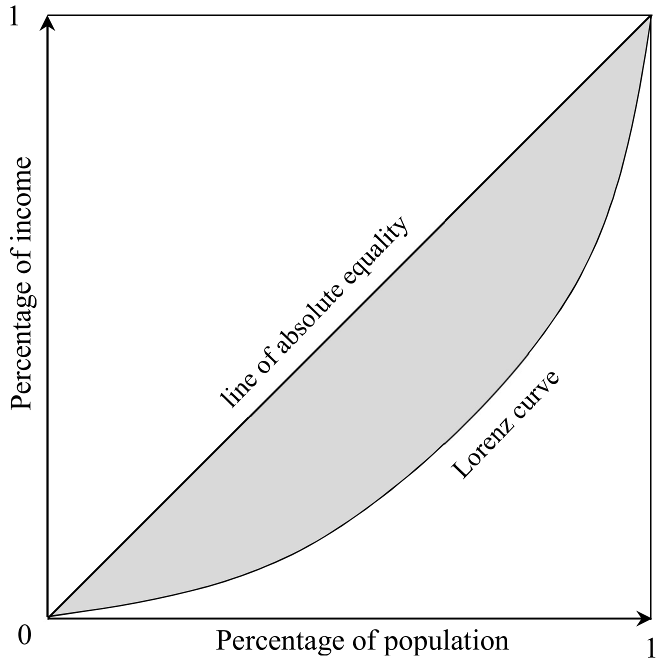
Gini index. In economics (and later in other fields, as well) a Gini coefficient (or index) is used to measure the magnitude of a program's equality (or inequality). This index is based on the Lorenz curve, which shows the percentage of income (or any resource) gained by any percentage of the population. Figure 3.3 helps to describe this curve and the associated Gini index (for more details, see e.g. Sen, 1973). The population (in our case, the investigators) are ordered by income (here grant money or number of papers) ascendingly on the horizontal axis in percent, while the income percentage is shown on the vertical axis. If everyone gains the same income (or grant money), the percentage of income gained by any percentage of population would be the same resulting in a diagonal line of absolute (perfect) equality. However, in reality, there is always some inequality in the distribution of resources and low-income people gain less income than their percentage in population. The locus of such real distribution points is called a Lorenz curve. Because of the existence of some inequality in resource distribution, the Lorenz curve always lies below the absolute equality line; and the lower the curve, the more unequal the distribution.

The Gini index is used to measure the magnitude of the inequality. Graphically, it is the ratio of the area between the equality line and Lorenz curve (shaded area in Figure 3.3) and the whole area of the triangle below the equality line. In practice, the population is divided into equal groups (like 10 groups or deciles) and their cumulative income percentage is calculated. When placed in a graph, these points give an approximation of the Lorenz curve. To estimate the Gini index, the following formula has been developed, based on the original version proposed by Dixon et al. (1987):

$$G = \frac{2\sum_{i=1}^{n} i x_i}{\sum_{i=1}^{n} x_i} - \frac{n+1}{n}$$

where n is the number of population groups indexed by *i* and x_i is the income share of the *i*th group. In this study, the population will be divided into 10 equal groups (deciles), with their grant money share represented by x_i.

Figure 3.3: Lorenz curve illustrated



Matthew Effect. Based on the interviews of some sociologist with Nobel laureates in the US and on his other experiences, Robert Merton (1968) developed the idea that famous scientists often get more credit than their comparatively unknown colleagues, for performing similar work. He called this phenomenon the Matthew effect,[1] which in his opinion goes beyond mere reputation and reaches the communication system that finally affects the allocation of scientific resources. Later (Merton, 1988), he expanded the concept and stated that advantages as well as disadvantages accumulated as a result of the working of this concept. In the context of this study, it would imply that those investigators who get some resources in the beginning and gain some advantage over others, gradually and continually will get more and more of the resources, which leads to an ever-widening gap among the investigators.

[1] The name was derived from Matthew's Gospel in New Testament, where there is a passage implying the same concept.

Although the presence of the Matthew effect can be described verbally, there are no precise tools to measure its magnitude. However, some measures can be applied to get a rough approximation of this factor. Since the investigator agents start simultaneously and they are assumed to be homogeneous, any differences emerging during the simulation horizon may be attributed to such Matthew effect. Apart from the Gini index described above, there is the concept of the power law or scaling correlation, which states that in some cases, there is a scaling correlation between two variables. Generally, a power law relationship exists between y and x when $y \propto x^{\alpha}$; α is called the scaling factor. To see if there is such a relationship, logarithms of y on x can be graphed against each other to see if the data lie on a line or not. In practice, a regression is run for $\ln y$ and $\ln x$ and the statistical significance of the relationship is checked. In the case of this study, it is hypothesized that such a relationship exists for the number of papers and the number of people having those numbers of papers; the same may hold for grant amounts and number of people. (Since Alfred J. Lotka was the first to study such a phenomenon, it is called Lotka's law defined as "the frequency distribution of the number of papers per author follows an inverse power law" (Watts and Gilbert, 2014, p. 138)).Put differently, it is speculated that only a small number of investigators publish many papers (or win large sums of grant money), while there are lots of others) with a small number of papers or grants.

3.4 Conclusion

The structure of the model is described in this chapter and must be translated into computer coding in Java language (in AnyLogic) for execution. There has been a lot of "back and forth" communication in the programming phase, which finally results in a working program for simulation. The computer model then is ready for finding bugs, for calibration adjustments and finally, for policy scenario simulations.

Chapter 4:

MODEL JUSTIFICATION AND VERIFICATION

In order to make sure that the model works in an acceptable manner, some tests are used to see if the main results are stable. The results of the simulation are summarized in a few measures discussed in the previous chapter: Papers per Million Dollars (PMD) for efficiency and Paper and Grant Gini Indices (PGI and GGI, respectively) for equity. The factors expected to affect the results are the values chosen for some parameters of the model and simulation. Sensitivity analysis is performed to check the impact of these factors and the results are reported below for each in turn; the baseline appears in bold in the tables.

4.1 A Brief Review of the Model[1]

Although the model and its technical features were described in the previous chapter, a brief review of the model follows, with the assumptions presented in Table 4.1. There are 600 homogeneous university investigators distributed across six regions in Canada. They simultaneously start their academic investigations at time 0, which take 2 years to be accomplished. Any investigation then results in 1 to 5 papers to be published in a period of 3 to 12 months.

The granting agency (GC) starts the grant competitions at the end of year 2 and then holds them every three years. Since there is no output for the investigators at the end of year 2, the grant approval for the first competition depends merely on chance and some investigators are randomly selected to receive grants. The grants are used to carry out investigations which yield some papers (with differing qualities and scores in later competitions), as mentioned above. Now those who win the grants by chance at the first competition, publish more papers which thereafter will be considered for grant proposal assessment. If the approval of the grants from the second competition and onward is only based on past paper performance, then the investigators who did not receive grants through the first competition would not stand a chance to win any grants in future; this reality justifies adding a random score besides the paper score. From that point onward, the competitions occur in the same way until

[1] The model can be found at: https://runthemodel.com/models/3353/

year 26, with those winning grants producing more papers. In cases where investigators have no grants, they fund one investigation by other sources.

Table 4.1: Assumptions of the model and parameter values

Parameter	Value	Considerations
Number of investigator agents	600 people	Scenario analysis
Budget per competition	\$50 million	Scenario analysis
Grant size	\$120-500 thousand	Scenario analysis; triangular distribution (120, 500, 240)
Competition gap	3 years	Scenario analysis
Simulation horizon	26 years	Sensitivity analysis
Investigation duration	24 months	Sensitivity analysis
Without grant annual fund	\$24000/ year	Sensitivity analysis
Papers per investigation	1-5	Sensitivity analysis; uniform distribution (1, 2, 3, 4, 5)
Publication duration	3-12 months	Triangular distribution (3, 12, 6)
Paper quality score	1-4	
Investigation cost	\$2000/ month	
Grant writing time	6 months	
Grant assessment time	1 month	

Source: Research findings

4.2 Model Justification

In order to have the model work in a reasonable way, some assumptions and adjustments are made. These assumptions are later analyzed to ensure that their variation does not significantly change the results. The first kind of assumption concerns regional distribution of the main agent (investigator) population; the second one pertains to the granting procedure; and the last one concerns the necessary number of simulation runs to make sure that model randomness does not bias the results.

4.2.1 Agent Population and Distribution

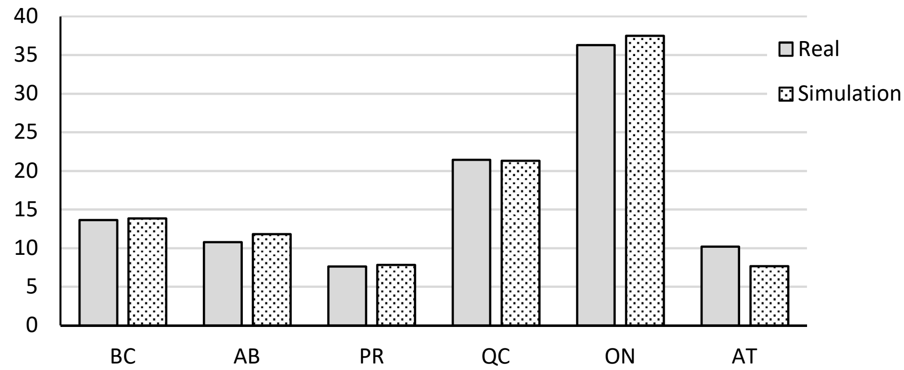
About 600 people participated in the ABC competition in 2009, whose robust number of participants became the basis for its choice as an investigator agents' population. To allocate this population to regions (consistent with six Genome Centres in British Columbia, Alberta, the Prairies, Ontario, Quebec, and the Atlantic region), the latest available statistics regarding full-time teaching staff in Canadian universities was utilized. The provincial distribution of university faculty members across Canada in the academic year of 2010-2011 (the latest available statistics) is presented in Table 4.2. Regrouping Manitoba and Saskatchewan as the Prairies (PR) and the four Eastern provinces as the Atlantic region (AT) gives the regional distribution in Figure 4.1. The third column of Table 4.2 is the real percentage distribution, which is used as a custom distribution in the model to assign a random region to investigators. As shown in the last column of Table 4.2 and in Figure 4.1, there is a small discrepancy between the real distribution and that of the random assignment by the model in some regions, but that is negligible here.

Table 4.2: Number of full-time teaching staff at Canadian universities in 2010-11 and agents' regional distribution

Province	Number of Teaching Staff	Real Regional Distribution (%)	Simulation Regional Distribution (%)
Newfoundland and Labrador	946	10.2	7.7
Prince Edward Island	247		
Nova Scotia	2,170		
New Brunswick	1,228		
Quebec	9,629	21.4	21.3
Ontario	16,307	36.3	37.5
Manitoba	1,776	7.6	7.8
Saskatchewan	1,660		
Alberta	4,846	10.8	11.8
British Columbia	6,125	13.6	13.8
Canada	**44,934**	**100**	**100**

Source: Statistics Canada (2016)

Figure 4.1: Distribution of university teaching staff in six regions in 2010-11



4.2.2 Granting Procedure

A couple of different ways were tried for grant approval. It seems logical that the approval should be based on past performance of the applicants, as reflected in their resumes. Therefore, the initial try was to use a probability distribution which translates the paper score into a probability value that is randomly chosen for granting; the higher the paper score, the higher the chance of approval. However, when the results were checked, it was surprising that some people would not receive any grants, regardless of their paper score. The reason for this is that the model starts deciding from the first person onward and, at some point, the budget is allocated completely without almost half the population having any chance to win a grant. Since this method is dependent on the budget and the budget itself is one of the policy variables, some other method is required. The next step is to make the model evaluate everybody before making the decision.

The alternative method is to have the model score all the investigators based on their paper output and rank them for fund allocation. This proves much better and is more realistic, but some extra diversity is needed. Since in reality papers do not have the same quality and relevance, paper scoring is added to the model. Based on the Australian Research Council's ERA initiative, which classifies journals into four groups or four tiers of quality (Wikipedia, 2016), the following quality distribution for the papers is considered: A^*, A, B, and C (1-4) with percentages as 5, 15, 30, and 50, respectively. In other words, the papers in the model are assumed to follow the above distribution for their quality and they are scored 1-4 accordingly. The scores for the papers are given randomly with the mentioned distribution upon being published and the paper score would be saved and accumulated for each investigator. Again, since there is no output for the first round of competitions, granting follows a random process until the next round, where the paper score ranking determines how funds are allocated. However, there is another problem which triggers the last remedy in the procedure.

Using the above procedure would result in a situation where most of the first-round grant winners would win the next competitions, leaving non-winners with no chance. Since in reality this is rare and the emphasis or target fields of the competitions vary, a random component is added to the model scoring, whereby every investigator gets a random score, ranging from zero to the highest paper score (a random number between zero and one is multiplied by the highest paper score). This random scoring restarts with every round. The total score, which is the sum of the paper and the random score, is used as the criteria for allocating funds; the investigators are organized in descending order, according to the total score; and the respective investigators are allocated the grants until the budget is spent.

Table 4.3 shows the results of procedures without and with random scores. When there is no random score and the grants are allocated merely using a paper score, almost all of the 171 investigators who win the first round (with a budget of $50 million) also win the next rounds, with 162 of them (95%) winning grants in all of the next competitions. On the other hand, the random score correction excludes most of the first-round winners in subsequent rounds, giving a chance to those who failed in the first and other rounds.

Table 4.3: Distribution of successive grant winners with and without including random score

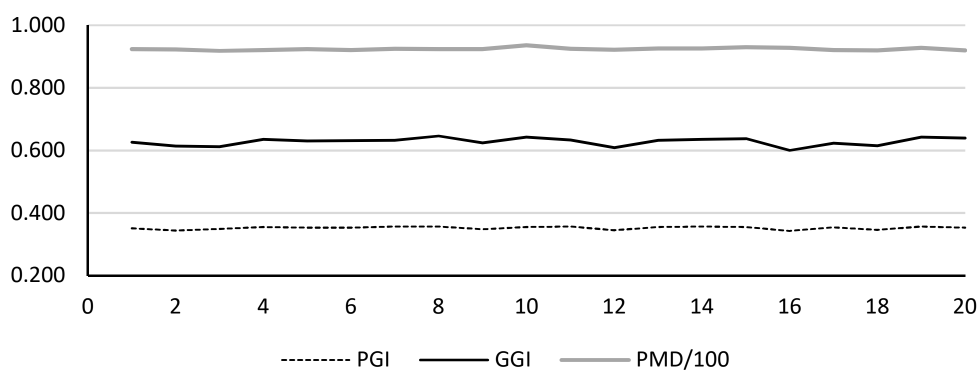
Number of grants won	**Without random score**		**With random score**	
Total	171	percent	171	percent
1	4	2.3	80	46.8
2	1	0.6	30	17.5
3	4	2.3	19	11.1
4	0	0.0	8	4.7
5	0	0.0	11	6.4
6	0	0.0	6	3.5
7	0	0.0	2	1.2
8	162	94.7	15	8.8

Source: Research findings

4.2.3 Randomness

One of the factors affecting the results in simulation models is the random number choice made by the model. In our model, some of the random processes include the assignment of region to investigators, choosing random investigators in the first round of granting, the random number of papers per investigation, and the random score. These random factors make the results differ for each run of the model. In order to see the magnitude of this factor, the model was run 20 times with different random seeds (i.e. the random number generator of the software); the resulting values for the three measures (PMD, PGI and GGI) appear in Figure 4.2. It is evident from the graph that the results do not change significantly and therefore there would be no need to run the model a couple of times and report the average as the final result.

Figure 4.2: Variation of efficiency and equity measures to randomness of the model



Some statistics regarding these 20 simulation runs are calculated and reported in Table 4.4. The maximum, minimum, and mean values for PMD, PGI, and GGI are shown in the table along with the standard deviation (a measure of variation or dispersion which is the average of the deviations from the mean) and coefficient of variation (CV which is standard deviation divided by mean). As expected, the standard deviation values and more importantly, the CV values are very small, implying the insignificance of the random factor. Therefore, the results below and the next chapter will be based on one single run of the model.

Table 4.4: Some statistics regarding the impact of model randomness on the results

Statistics	PMD	PGI	GGI
Minimum	91.8	0.34	0.60
Maximum	93.6	0.36	0.65
Mean	92.4	0.35	0.63
Standard Deviation	0.401	0.005	0.012
Coefficient of Variation	0.434	1.31	1.96

Source: Research findings

4.3 Model Verification

There are some parameters in the model for which some values have been assumed. However, their values can be different in reality. Therefore, some other values can be attributed to test if the results have acceptable stability. Additionally, the way the model responds to the parameter changes is important for ensuring that the model works properly. Below, some sensitivity analyses are performed to test the model's behaviour. In running the simulations, every time only one parameter is changed to check its impact, while the others remain fixed at the baseline level.

4.3.1 Papers per Investigation

It is assumed that every investigation results in at least 1 and at most 5 papers. A uniform probability distribution is used to pick a number from 1 to 5 (with the same chance) as the number of papers when the investigation is completed. As was predicted, when the maximum number of papers per investigation is reduced to 4 or 3, the efficiency measure of the simulation decreases accordingly, since the whole output of the system goes down (see Table 4.5). However, the changes in the equity measures are not significant and the choice of the current value for this parameter does not have a considerable impact on model output; as with PMD, since the focus of the study is on the comparison of the results from policy variable changes, it does not make a difference in the comparisons.

Table 4.5: Sensitivity analysis results for the maximum number of papers per investigation

Maximum Number of Publications	PMD	PGI	GGI
3	61.2	0.35	0.62
4	76.4	0.35	0.61
5	**92.2**	**0.35**	**0.61**

Source: Simulation runs

4.3.2 Investigator Agent Population

As mentioned above, the target population size is assumed to be 600, based on the number of applicants for the 2009 competition for GC grants. Although the number of potential applicants (or real target group) should have been greater (some people did not apply for the grants), this number seems right because in the model all of the investigators enter all of the competitions. However, a sensitivity analysis is conducted to see the impact of population size on the results. Although the grant money is fixed at $50 million, the PMD value changes as a result of investigations and papers which are published without grants. There is no clear trend in PGI but GGI increases with population size, since a smaller subgroup of the population can receive grant money when the whole group gets larger and this worsens the money distribution among the investigators. Since the population size can be looked at as a policy variable from a different perspective, these changes will be discussed in more detail in the next chapter.

Table 4.6: Sensitivity analysis results for the population size

Population Size	PMD	PGI	GGI
300	66.2	0.31	0.41
600	**92.2**	**0.35**	**0.61**
1200	146.2	0.33	0.82

Source: Simulation runs

4.3.3 Simulation Horizon

The simulation horizon was initially considered to be 25 years, but another year was added later in order for the model to include the results of the last competitions. The first competition takes place at the end of year 2 and then goes on by 3 years giving the last round (number 8) at year 23. The investigations funded by these last grants are concluded at year 25 and the papers resulting from it are published the following year. The papers coming from non-GC funds accumulate over time, leading to a constant increase in PMD over time, as is clear from Table 4.7. With a longer horizon and more competitions, the chance for every investigator to receive a grant and publish more papers increases, resulting in a better distribution of papers and grant money. Again the choice of time horizon for a public program and the number of granting rounds might be a policy issue; the policy implication is that longer-term granting is better on the grounds of both efficiency and equity.

Table 4.7: Sensitivity analysis results for the simulation horizon

Number of Competitions (Time Length)	PMD	PGI	GGI
4 (168 months)	79.6	0.41	0.68
6 (240 months)	85.5	0.38	0.63
8 (312 months)	**92.2**	**0.35**	**0.61**

Source: Simulation runs

4.3.4 Investigation Duration

The investigations are assumed to last for two years, prior to giving birth to papers. This is the average period required to complete for a Master's thesis and sounds reasonable. Changing this period to one year results in a large increase in PMD, since the monthly cost of the research stays the same. Therefore, not only does the number of papers produced from the grant money double, but the number of papers produced from non-GC grant investigations also doubles, leading to a more than 100 percent rise in PMD (see Table 4.8). On the other hand, the equity measures worsen slightly, which is interesting and may reflect in the result of a change in paper distribution which in turn influences the money distribution (chance of winning grants). Nonetheless, the model's behaviour is normal.

Table 4.8: Sensitivity analysis results for the investigation duration

Investigation Duration (months)	PMD	PGI	GGI
12	208.0	0.36	0.63
24	**92.2**	**0.35**	**0.61**

Source: Simulation runs

4.3.5 Annual Funds with No Grant

One of the assumptions is that when investigators do not succeed in getting grants, they can collect funds from other sources that are large enough to run one investigation. With any one investigation taking two years and costing \$2000 per month, the annual no-grant fund will be \$24000. Reducing this sum by half should result in a decrease in PMD (no-grant investigations and papers are cut by half) and a rise in PGI, since the investigators failing in grant competitions then produce fewer papers. However, the increase in GGI is almost unexpected (compared with Table 4.8), since nothing directly happens to the granting procedure. It seems that the reduced paper output of non-winners decreases their chance of winning grants in subsequent rounds, thus making the grant distribution more inequitable (see Table 4.9).

Table 4.9. Sensitivity analysis results for the annual no-grant fund

Annual Fund ($1000)	PMD	PGI	GGI
12	71.8	0.48	0.66
24	**92.2**	**0.35**	**0.61**

Source: Simulation runs

4.4 Conclusion

The model was tested according to parameter changes and real-world logic and was shown to be working well. The next step will be to interpret the simulation results and to infer some implications for policy makers— these are addressed in the following chapter.

Chapter 5:

RESULTS AND DISCUSSION

Having tested the validity of the model and its behaviour, the results from simulations will here be presented and discussed, using the values of the parameters discussed in the previous chapter. First, the results of the baseline scenario are presented and discussed, followed by the results of some alternative policy scenarios.

5.1 Baseline Scenario Results

The baseline parameter values are presented in Chapter 4, along with the basic values for policy variables. Using those benchmark values for all of the parameters and policy variables, the simulation gives the baseline results to be discussed here. As before, the three measures of PMD, PGI, and GGI will be used to analyze outcomes. However, another measure will be used to analyze equity, so as to capture a phenomenon called the "Matthew Effect," which will be explained below.

5.1.1 Baseline Scenario Results and Regional Differences

The total paper output of the system is about 37,000, which is produced in 26 years of the model simulation. The trend of this output is shown in Figure 5.1, along with the number of investigations. Since papers come out of investigations, their behaviours are similar (paper graph follows the investigation one) and because of the simultaneity between initiation and conclusion of investigations, which are triggered mostly by grant rounds, the graphs have some kinks. At the end of the simulation horizon, a total of 12,309 investigations have been accomplished, resulting in an average of almost three papers per investigation, as expected.

As defined in the previous chapter, dividing the number of papers by the grant money spent results in the efficiency measure of PMD. This measure can be calculated at any time and the graph is seen in Figure 5.2. With 50 million dollars spent at every competition and having eight competitions, the total amount of money spent would be $400 million; in order not to produce any negative balance, the model continues to extend grants, as long as the balance is equal to or greater than $500,000 and that is why $229,000 is left and the grant sum is $399.8 million. Because no papers were produced in the initial years and investigation outcomes were the sudden publication of the investigation outcomes, the graph rises steeply and then flattens gradually, since the numerator (papers) and denominator (grant dollars) behave as before and the initial delay wears out. However, although the money is spent quickly at any round, the publication process is more gradual (3 to 12 months for a paper) and that is the cause of non-smoothness of the trend. At the end of the simulation horizon (month 312) the PMD measure reaches 92.2, meaning that every million dollar of grant money results in an average of more than 90 papers. It should be remembered that some of the papers originate in investigations carried out from non-GC money.

Figure 5.1: Trend of investigations and papers during the simulation horizon

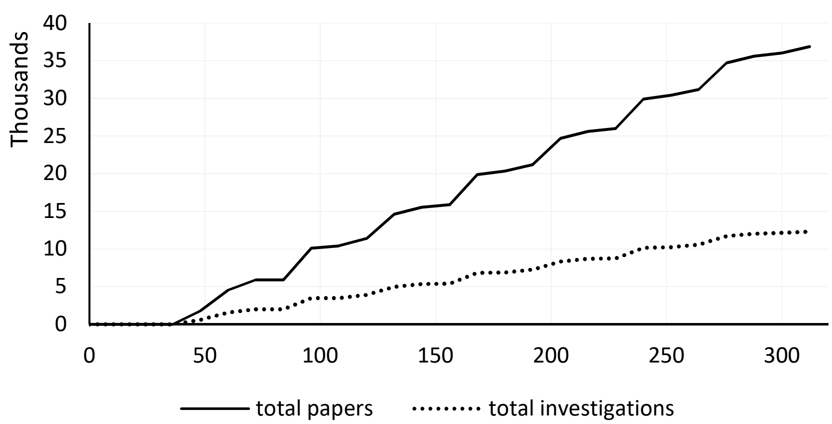
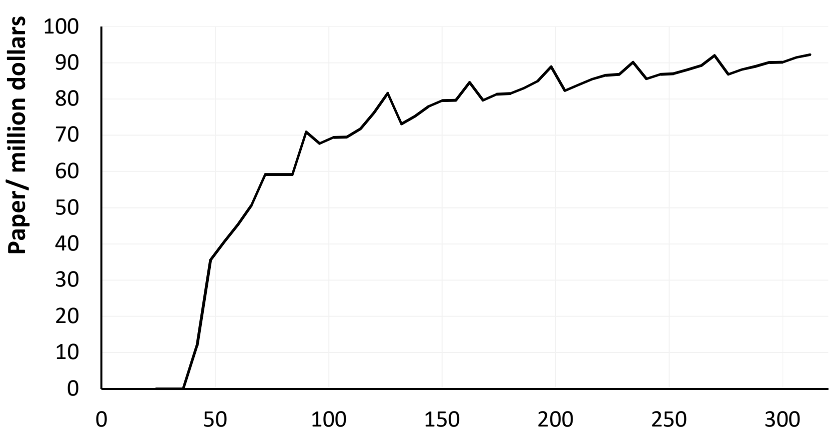


Figure 5.2: Trend of PMD



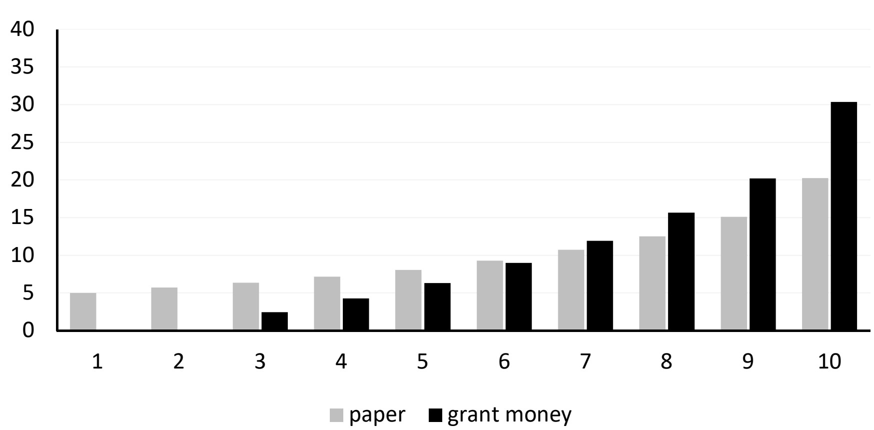
The equity deals with the distribution of resources and the outcomes among the investigators. Table 5.1 shows the distribution of the papers and grant money among 10 ascending equal groups; investigators are ordered according to their number of paper publications or the grant money they have won. And the sum of the values for every 10 percent of them (60 people) are considered as deciles. According to the table, the first decile publishes just 5 percent of the total papers but the tenth decile authors 20%. In other, words, the top decile of investigators publish 4 times as much as the bottom 10 percent. A glance at Table 5.1 and Figure 5.3 reveals that the distribution of grant money is much worse: the bottom 20 percent receive nothing, while the top decile receives more than 30 percent of the total grant dollars. The difference originates from the assumption that people can investigate and publish even without grants.

Table 5.1: Paper output of the system and its distribution among investigators

Decile	Papers			Grant Money		
	Number	%	Cumulative %	$ million	%	Cumulative %
1	1,839	5.0	5.0	0.0	0.0	0.0
2	2,105	5.7	10.7	0.0	0.0	0.0
3	2,329	6.3	17.0	9.7	2.4	2.4
4	2,627	7.1	24.1	17.0	4.3	6.7
5	2,965	8.0	32.2	25.2	6.3	13.0
6	3,420	9.3	41.5	35.8	9.0	21.9
7	3,955	10.7	52.2	47.5	11.9	33.8
8	4,612	12.5	64.7	62.4	15.6	49.5
9	5,569	15.1	79.8	80.7	20.2	69.6
10	7,451	20.2	100.0	121.3	30.4	100.0
Sum	**36,872**	**100.0**	**-**	**399.8**	**100.0**	**-**

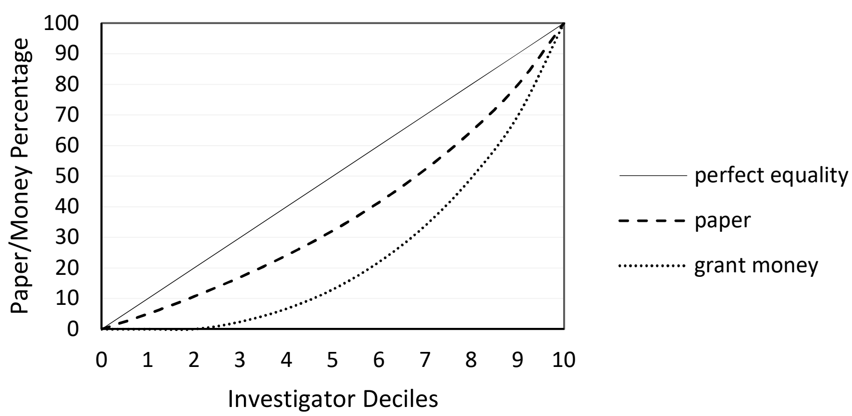
Source: Research findings

Figure 5.3: Distribution of papers and grant money among investigator deciles



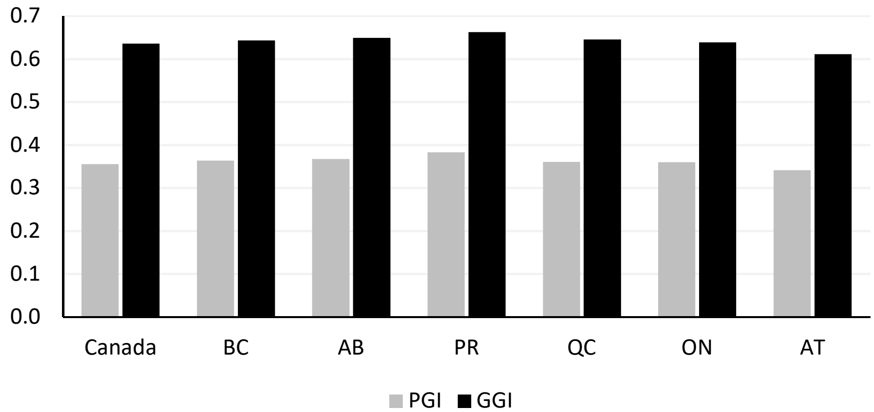
The equity is better measured by the Gini index which is based on the Lorenz curve (cumulative distribution). From the cumulative distributions in Table 5.1, the Lorenz curves for paper and grant money are drawn in Figure 5.4. It can be noticed that there is some inequality in the distribution of papers and grant money (both curves are lower than the 45-degree line of perfect equality) but the grant money curve is much lower than the distribution of paper curve. The Gini indices computed for papers (PGI) and grant money (GGI) are 0.346 and 0.606, respectively; the closer the value to zero, the more equal the distribution.

Figure 5.4: Cumulative distribution of papers and grant money (Lorenz curves)



The regional attribution of investigators (Figure 5.5) is quite random and there seems to be no reason for regional disparity in terms of outcome measures. However, since the size of the regions are different, there may exist a size effect which was checked in the model. There is quite high probability of finding a significant difference among regions in a single run (e.g. in one of the runs PMD varies from 86.8 to 96.6 with the national value being 92.7). Therefore, five different runs were carried out the average of the values represented in the graph of Figure 5.5; PMD values were 92.7, 94.3, 91.7, 92.8, 93.3, 92.7, and 95.9, respectively. The differences are not significant and it can be inferred that they are random.

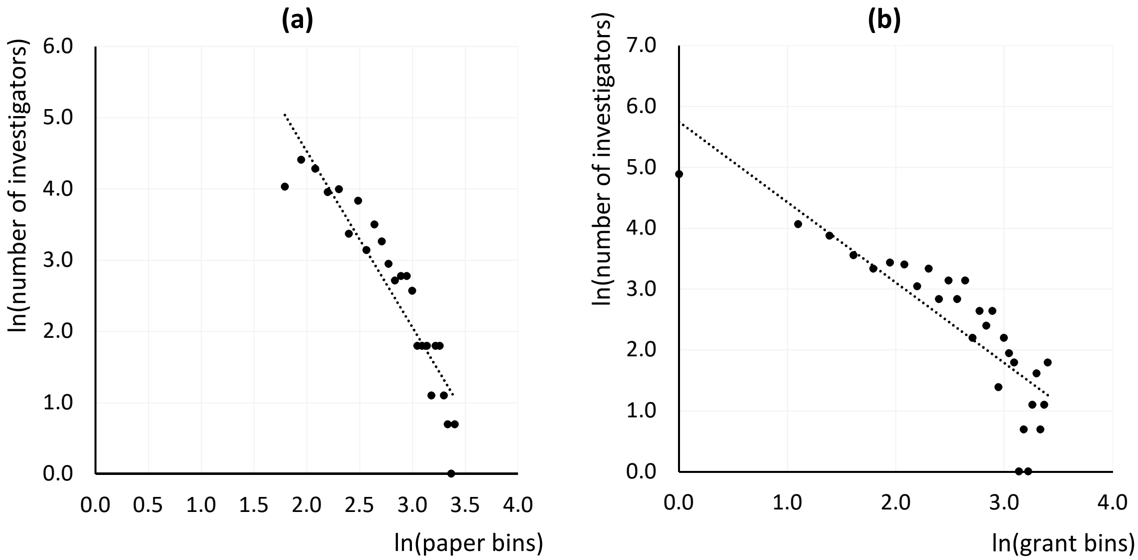
Figure 5.5: Equity measures in terms of regions



5.1.2 The Matthew Effect and Power Law Relationships

As another measure of equity, power relationships between both number of papers and grant money on the one hand, and the number of investigators publishing the papers or winning the grants on the other, were used to examine the concentration of the publications among the investigators. The number of papers varied from 24 to 166 and the money granted varied from 0 to $2.7 million and that is why they were categorized into 30 equal bins. The bins were numbered from 1 to 30 which is considered *x*, and their contents (the number of investigators in every bin) were counted to represent *y*. Taking logarithms of these variables and presenting them in a graph would result in Figure 5.6 for papers (a) and grant money (b). The dots represent real values and the dotted line is the line fitted on the data; since some of the bins were empty, the number of dots (observations) is less than 30.

Figure 5.6: Power law relationship for papers (a) and grant money (b)



Obviously, there is a near-linear relationship in both cases in Figure 5.6. However, to check the statistical significance of the relations and determine the scaling factor (α), a regression had to be run over the data. This was carried out in Excel using its Data Analysis feature and the results are shown in Table 5.2. The equations are shown in the second column along with the *t* statistics. All the parameters (constants and slope (α) parameters) are significantly different from zero, meaning that there is a significant relationship between the variables. The number of observations used (n), the coefficient of explanation (R^2), and the overall significance of the regression (*F* statistic) are reported in the last column. Again the regressions are quite significant and most of the changes in dependent variables are explained by the *x* variables (or there is a high correlation between the variables). In summary, it means that the power law holds in both paper and grant money cases.

Table 5.2: Results of the log-linear regression estimations for power law relationships

	Estimated Relationship	Statistics
Paper analysis	$\ln(y) = 9.5 - 2.47\ln(x)$ *t* values 16.1 11.9	n=25, R^2=0.86, F=141.8
Grant analysis	$\ln(y) = 5.7 - 1.32\ln(x)$ *t* values 14.3 8.7	n=29, R^2=0.74, F=76.5

Source: Research findings

5.2 Alternative Policy Scenarios

There are some tools that granting agencies can use to influence the outcomes of the model in the long run. However, the outcomes are measured from two perspectives (efficiency and equity) which sometimes do not behave similarly and between which there is a trade-off. The variables that can be manipulated by a granting agency such as GC are the total sum of money allocated in any competition, the size of the individual grants, the size of the target group, and the gap between the two consequent competitions (or number of competitions). The impact of changes in these variables is explored below, followed by an analysis of their combinations. Table 5.3 gives a summary of the scenarios to be discussed; the instruments will be changed one by one and the combination of tools is left for the next section.

Table 5.3: Summary of the single-instrument policy scenarios

Scenario Variable (Policy Instrument)	Unit	Values
Competition budget	$ million	50, 100, 150
Grant size (min, mode, max)	$ thousand	(120-240-500), (120-360-500) (120-240-1000), (120-480-1000)
Target group size	persons	600, 400, 200
Competition gap	years	3, 4, 5

Source: Research findings

5.2.1 Competition Budget

It was assumed in the benchmark model that the granting body allocates $50 million for every competition. This would result in a PMD of 92.2 and PGI and GGI values of 0.35 and 0.61, respectively. With an increase in the amount of the budget, the efficiency drops and equity improves. However, the magnitude of changes is not the same: as seen in Table 5.4, a 100 percent increase in the budget results in a 28% drop in PMD, a 33 percent decline in GGI and an only 10 percent drop in PGI. Therefore, increasing the budget does not seem to be a good policy, due not only to the decrease in paper productivity, but also to the small improvement in equity indices. It should be noted that with a $50 million budget and the grant sizes assumed, about 175 people (29%) receive grants in each competition. It is interesting that another $50 million raise in the budget (from 100 to 150), which is equal to 50 percent, leads to a 13% decrease in PMD (almost half the previous amount) and a 27 and 41 percent decrease in PGI and GGI, respectively. These figures show that the measures do not change linearly with the budget.

To see the behaviour of the measures in response to budget change, different values were tried for the budget (10 to 150 million dollars by increments of 10). The outcome measures are drawn as graphs in Figure 5.7 with equity measures on the left axis and PMD on the right. It is evident that PMD decreases with a declining slope (like a power function), meaning that its response to budget increases declines. The same kind of response is shown by GGI but with a smaller change in its slope. The most interesting behaviour is that of PGI which rises to $60 million and then starts to decline. Smaller budgets are not favorable due to the great inequity in grant money and, since large budgets are impossible to raise, leaving the mid-range of the budgets for choice. However, other factors remain to be considered before one makes such a choice.

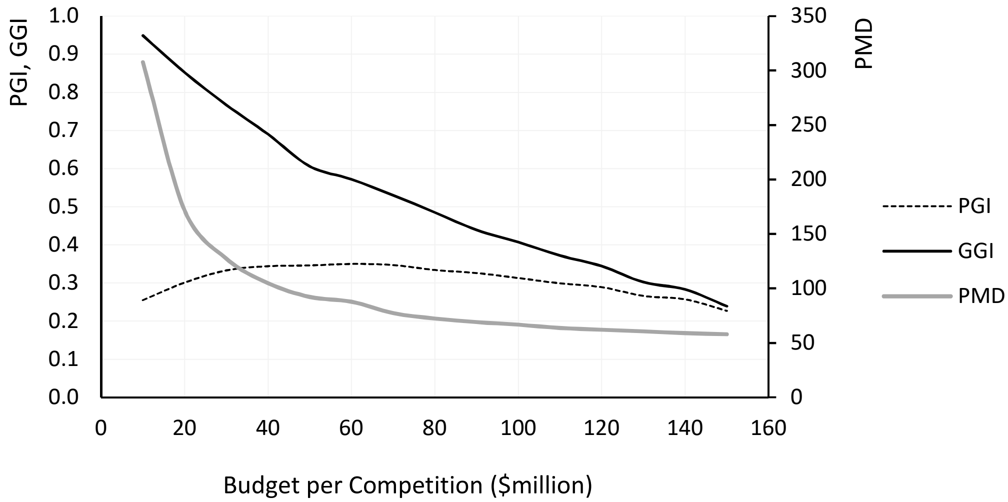
In analyzing the variable of the budget, some consideration should be given towards the administrative costs of any grant program or agency and also the opportunity costs of the participants. Part of the costs of handling the competition and granting process is fixed and does not change with the amount of budget allocated, implying that their average per dollar granted declines with an increase in the budget. Similar reasoning governs participant opportunity costs, where it does not make sense to hold a competition with 600 participants, in which only a small percentage of applicants are approved. It can be argued that to better manage participant opportunity costs, narrowing the competition and decreasing the size of the target group might be better.

Table 5.4: Effect of increasing the grant budget on outcome measures

Budget Allocated for Each Competition	PMD	PGI	GGI
50	**92.2**	**0.35**	**0.61**
100	66.7	0.31	0.41
150	57.9	0.23	0.24

Source: Simulation runs

Figure 5.7: Behaviour of outcome measures to changes in grant budget



5.2.2 Grant Size

The grant size is assumed to vary following a triangular distribution, with a minimum of $120,000 and a maximum of $500,000 and a mode of $240,000. The results for four different combinations of these values are shown in Table 5.5. Increasing the mode has resulted in a very small improvement in PMD but a small worsening of equity measures. Doubling the maximum mode results in a worse distribution of papers and money. Doubling both the maximum and the mode of the triangular distribution above, gives the worst outcome with a 36 percent rise in the money Gini index. Again there are administration and opportunity costs issues, so that it is not easy to decide what range of grant size is the best.

Table 5.5: Effect of increasing the grant size on outcome measures

Grant Size	PMD	PGI	GGI
120-240-500	**92.2**	**0.35**	**0.61**
120-360-500	93.5	0.38	0.69
120-240-1000	97.8	0.42	0.74
120-480-1000	99.7	0.46	0.83

Source: Simulation runs

5.2.3 Target Group Size

One of the assumptions in the model was the openness of competition to the whole population of investigators. However, there are occasions where the competitions are open only to some specific disciplines or where the focus of the competition itself is on some specialized fields. Such a situation means that not everybody is eligible to apply for the grants, leading to a decline in target group size. Table 5.6 reports the outcomes for two different target group sizes of 400 and 200. Although these calculations can be done for a specific subgroup of the population, like those in specific fields, due to the homogeneity of the investigator agents, it amounts to a smaller population size here. Therefore, the outcome is expected to be similar to that of changing the budget amount —a budget decrease works in the same way as a population increase— and the results are comparable to Table 5.4.

Table 5.6: Effect of limiting the target group on outcome measures

Number of Investigators per Competition	PMD	PGI	GGI
600	**92.2**	**0.35**	**0.61**
400	74.6	0.35	0.51
200	57.4	0.22	0.24

Source: Simulation runs

5.2.4 Competition Gap

There is another variable the granting agency can alter to influence the outcomes and that is the time period between two competitions. In the baseline model, competitions are held every 3 years with a total of 8 competitions and a total budget of $400 million allocated. Raising the gap in time to 4 years would mean that there would be a total of 6 competitions (starting from year 2 and ending in year 22) with a budget of $300 million. If the gap is raised to 5 years, there would be 4 rounds with a budget of $200 million. According to Table 5.7, this kind of policy improves the efficiency but has a small negative impact on money distribution. Provided that the same grant budget ($400 million) should be allocated, there would be $100 million for every round resulting in values of 89.3, 0.30, and 0.43 for PMD, PGI, and GGI, respectively. When compared with results of the base scenario reported in the first row of the following table, it is evident that with a small loss in efficiency, there would be a large win in equity.

Table 5.7: Effect of increasing the competition gap on outcome measures

Competition Gap (years)	PMD	PGI	GGI
3	**92.2**	**0.35**	**0.61**
4	116.9	0.34	0.65
5	131.7	0.33	0.67

Source: Simulation runs

5.2.5 Optimal Scenario

Because of the presence of two competing criteria for evaluation of programs (efficiency vs. equity) it is hard to find the optimal level of policy instruments. A change in a policy variable such as budget allocated for each round of competitions, which is discussed below, would change both PMD and GGI in different directions, making it impossible to tell what direction of change is better. However, provided that there is some knowledge of the weight given by policy makers (hence to society) to each of the criteria (efficiency and equity), further guidelines can be given regarding the optimal scenario. The weights can be used to combine the two measures to get a single criterion with which the optimal level of a policy tool can be approximated. Table 5.8 gives the simulation results for different values of grant budget starting from $10 million and increasing by 10 to a maximum of $150 million.

As predictable from Figure 5.7, none of the outcome measures change in a linear manner: PMD changes substantially first but slows down gradually; GGI starts with moderate changes followed by smaller changes and larger ones at last. It can be noticed that with lower budgets, a great amount of efficiency should be sacrificed to get a moderate gain in equity (e.g. 44 percent reduction in PMD versus 10 percent gain in equity). However, at a budget of 50, the changes are almost the same (12.0 versus 12.2). Moving from a budget of 50 to 60 makes sense, since a loss of 4.8% in PMD leads to a gain of 5.6% in equity. The next move from a budget of 60 to 70 is not good, but there are much better moves down the way. Dividing the two percentage changes gives a measure called elasticity (of changes in PMD with respect to changes in GGI) which are reported in the last column. A unit elasticity means that 10 percent change in PMD coincides with the same percentage change in GGI. Higher budgets result in smaller improvements in PMD but great gains in equity, making the elasticity very low. Nevertheless, large budgets are not feasible and small budgets are unjustifiable in terms of the administrative costs involved; a budget range in the middle is plausible. The optimal levels should be sought in such a range and if the two criteria are of the same weight, the neighbourhood of an elasticity value of one (unit elasticity) would be optimal.

As noted above, finding the optimal levels of policy tools requires further information that lies beyond the scope of this study. On the one hand, knowledge of the administrative costs of granting is necessary and on the other hand, the comparative values of outcome criteria are needed. There are methods to approximate the weights attached to each criterion (here, efficiency and equity) by policy makers and to use those weights to combine the measures and compute a single index. The optimal level of any policy instrument would depend on the comparative emphasis a society puts on either criterion.

Table 5.8: Impact of budget increase on efficiency and equity measures

Competition Budget (million dollars)	**PMD**	**GGI**	**% Change in PMD**	**% Change in GGI**	**Elasticity** (%ΔPMD/%ΔGGI)
10	307.7	0.95	-	-	-
20	171.6	0.85	-44.2	-10.2	4.33
30	127.3	0.77	-25.8	-10.0	2.59
40	104.8	0.69	-17.7	-10.0	1.76
50	92.2	0.61	-12.0	-12.2	0.99
60	87.8	0.57	-4.8	-5.6	0.85
70	77.3	0.53	-12.0	-7.3	1.63
80	72.4	0.49	-6.3	-8.5	0.75
90	69.1	0.44	-4.6	-9.5	0.48
100	66.7	0.41	-3.5	-7.3	0.48
110	63.7	0.37	-4.5	-8.6	0.52
120	62.1	0.34	-2.5	-7.5	0.33
130	60.6	0.30	-2.5	-12.2	0.20
140	59.0	0.28	-2.5	-6.3	0.39
150	57.9	0.24	-1.9	-15.5	0.13

Source: Research findings

5.3 Combination of Policy Tools

The policy alternatives discussed above dealt with manipulating single instruments. In practice, policy-makers are able to choose policy combinations consisting of multiple tools. Although a lot of combinations might be available for them with various values chosen for every instrument, only a few combinations are analyzed here, as an example. Suppose that the long-term budget of the granting body is $400 million, which must be allocated in grants over a period of 20 years. According to our model, it can start at year 2 and continue until year 22, in a couple of ways, as shown in Table 5.9. The first method is the baseline model in which the grand budget is allocated in 8 rounds of $50 million each, with grant size ranging from 120 to 500 (following a triangular distribution). The outcome measures have been discussed before. Now an alternative method with the same grant budget may be used, in which competitions are held every five years (4 rounds), with the same options for grant size. As seen before, such a policy change results in a small loss (3%) in efficiency but in a large gain in equity (i.e. 30 percent decrease in inequity index). Such an outcome could be achieved with holding the gap and round budget constant while cutting the grant size. According to Table 5.9, a grant mix of (min=100, mode=150, max=300) would lead to a GGI of 0.45 which is close to the previous case, but a smaller PMD (a 10 percent loss in efficiency, but a 26 percent gain in equity, in comparison to the baseline model).

A second improvement in the results can be achieved by combining the two changes: raising the gap (along with budget size) and reducing the grant size. As seen in the following table, now only a 12 percent sacrifice in papers a 64 percent gain in equity would be possible. Of course, there are other issues, like administrative limitations for the granting agency and other problems that have not been taken into account in our model. Nonetheless, the model gives some basic guidelines which can be utilized to set policy actions on a better course.

Table 5.9: Outcomes of some policy instrument combinations

Budget	Round Gap	Grant Size	PMD	PGI	GGI
50	**3**	**120-240-500**	**92.2**	**0.35**	**0.61**
100	5	120-240-500	89.3	0.30	0.43
50	3	100-150-300	83.3	0.27	0.45
100	5	100-150-300	81.0	0.19	0.22

Source: Simulation runs

5.4 Concluding Remarks

Having checked the validity and functionality of the simulation model in Chapter 4, the results were presented and discussed in this chapter. Depending on the relative importance of efficiency or equity for the policy makers, different tools and combinations of tools can be utilized to achieve specific goals. However, as usual with most systems, there are trade-offs between efficiency and equity: any improvement in one of them usually comes by compromising the other. While the model allows basic guidelines to be drawn from it, detailed and more precise policy recommendations are not possible, due to the limitations of the model. In the following chapter, some of these limitations will be discussed.

Chapter 6:

SUMMARY AND CONCLUSION

6.1 Summary

As a player in the Canadian innovation system, Genome Canada has been funding genomic research for almost two decades. It has held some competitions in the past to collect research proposals and approve some of them for co-funding. The continuation of such a program shows that policy-makers have been satisfied by its outcomes. However, any program can be improved to better serve society, and this study is an attempt toward that end. Its objective is to investigate the process by which the grant money can be turned into publishable knowledge and the impact of the tools available to influence that process.

To achieve the above objective, agent-based modelling (ABM) simulation technique was employed using AnyLogic. A model of the academic investigation process was built, in which investigators as agents turned their input (their own time and effort and funds) into an output (here, papers). The investigation itself was another type of agent in the model, along with the grant agent, with some assumptions in each case. A total of 600 investigators were active in the system, carrying out research and competing for the grants. In order to be able to compare the results of different policy courses or parameter changes, some measures for the outcome of the model were needed. Two types of such measures were defined, one of them capturing efficiency (number of papers per money granted) and the other one capturing the equity effects. For this last type, an index of the distribution of resources (grant money) and of output (papers which are resources to gain more grants) was used, which was originally developed in development economics. The Gini index showed how equally (or unequally) an endowment was distributed among the associated population.

Assuming values for the parameters of the model and its simulation, the model was used to obtain the index values. To check the functioning and validity of the model, some tests were carried out to make sure that it worked properly and to check if the results were stable enough in response to some parameter changes. After the procedures and the values of parameters were adjusted, the model was ready for policy simulation runs. The results revealed that the instruments of allocated budget per competition, the gap between competitions, the sum granted for any proposal, and the size of the target group may be utilized by a granting body to improve the efficiency and equity of the system. However, there is usually a trade-off between these two objectives and a loss in one of them is necessary to achieve a gain in the other. The tools can be combined in order to secure better results, but there are other factors that should be taken into account in making decisions.

6.2 Originality and Contributions

This study is original and innovative in three respects. First, this is the first attempt to apply agent-based modelling to the Canadian innovation system and simulate the impact of policy tools. To do this, a stylized generic model was developed and tested, using Canadian Genome Canada research funding as the illustrative context. Second, Gini coefficients were borrowed from economics and introduced here to check the equity aspect of public research funding. As far as I know, it is original to this study. Third, the Matthew effect and power law were studied and tested in a new context. Again, this is a unique contribution of this thesis.

6.3 Policy Implications

There is no agreed-upon solution for the efficiency-equity debate. From one viewpoint efficiency is favourable because it means a bigger cake for the whole society, which in the long-run can benefit everybody. But every country has some notable scholars and scientific stars who are supported to be able to obtain international credit and resources, in order to enhance reputation and leverage research impact. In the international arena, no single country operates to improve equity among countries but instead most seek their own benefits. From this respect, some equity is sacrificed in favour of efficiency to promote science stars; there can be no prominent stars if the equity goal is sought by allocating resources and outputs equally among the academia. Complicating this is that inherent differences in capabilities and effort can amplify unequal allocations of resources.

On the other hand, sometimes equity is used to promote the entrance of new people into the academy and, optimally, as emerging stars. Although efficient people may be able to turn scarce resources into a bigger output, those who begin as less efficient producers may be able to become more productive in future if given some support in the short run. The path of building new stars in effect is based on more equitable distributions. At another level, one might argue that public funds should always be used to promote equity—in that sense that supports the position for equal allocation of resources irrespective of the capabilities and differences of the people involved.

In our model, all the investigators enter the research business at the same time and one might suppose that the main objective should be efficiency. However, the first competition round starts with mere chance and the following rounds are to some extent dependent upon the performances in the previous rounds. But both chance and individual performance also matter. It is almost impossible to know who will get the best results with some inputs; better performance in the past does not guarantee efficiency in the future. People are different and respond differently to incentives and opportunities. Therefore, while a key criterion in research funding should be efficiency it is important that policy makers keep an eye on equity, as well. Path dependence that disenfranchise people completely undoubtedly undercuts efficiency.

In theory, it is possible to establish the relative importance of efficiency and equity using expert opinions, effectively combining them. A range of methods can be used to determine the weights for each objective which can be used to combine the two criteria into one. Such a method may settle the debate but it should be kept in mind that opinions change with the information provided for people and the way the problems are framed.

Based on the above arguments and the tools discussed in the previous chapter, a number of implications follow for policy makers. Collecting and allocating large sums of money for every competition means that more researchers will get the chance to obtain a grant. Provided that studies support the idea that the efficiency of a dollar spent in academic investigation is higher than the other sectors of the economy, such a policy seems justifiable, but if budgets are limited and not all research is equally efficient, small competition budgets with small grants for the participants may be preferable. Unless there is supporting evidence in favour of allocating funds to some specific fields, it is not recommended to hold specialized competitions which limit the applicants to some areas and disciplines. Decreasing the size of research target groups leads to a cut in efficiency in spite of promoting the equity among the members of the targeted fields. One strategy might be to increase the gaps between research programs, which could improve efficiency with a small loss of equity; there is also some evidence that larger gaps may be associated with lower administration costs. Consideration of other factors may change the above arguments but it will not be significant; some of such factors are discussed below.

6.4 Concluding Remarks

The agent-based model developed for this study is a simple one, assuming homogeneous investigators, applying for grants individually, and spending the grant money to carry out investigations and produce papers. Although some lessons can be learned from such a simple model, making it applicable to policy making and to real-world issues, other factors need to be included in the model. Below are listed some important aspects that are neglected in the model for the sake of feasibility. Since research collaboration is very important and its neglect may have profound effects on the results and some attempt had been made to include it in the model, it is treated separately below.

1) The investigators are assumed to be identical. In reality, there are different disciplines and fields working on genomics-related research, in which the research methods and tools will vary. Some disciplines need a lot of laboratory equipment and materials; yet others are labour-intensive; the social sciences are fundamentally different from natural sciences and there are differences between natural science disciplines, themselves. There may be differences in the skillsets of faculty members within the same department —for e.g., some people are research-oriented, while others prefer teaching. Such differences may and should be considered to develop a more realistic model.

2) The process of investigation may be changed to include various factors used in research (like equipment, materials, human services), as well as multiple outputs. The final output of investigations is not just papers but also patents (inventions) and training, as well. People learn by doing and such knowledge accumulates over time, leading to enhanced productivity for the future. Such features can also be added to the basic model used here.

3) The granting procedure and the mechanisms and expenses within the granting agency can be modelled and added to the model. Here the administrative costs of the granting body were neglected, along with the co-funding mechanism and its procedure.

6.5 Expanding the Model to Include Collaborations

Investigations are usually carried out via collaborations. There are cases where some investigator works alone or a graduate student is involved both helping and getting training. However, funded research is increasingly done through collaborations. In such cases, there is a principal investigator (PI) who initiates the process of application and leads a team of investigators. The idea for the research and application is born somehow and the PI starts gathering a team based on his network.

Every academic has a professional network where he/she knows some people through a range of communities (Sharma 2012 examined these networks in the context of the Genome Canada ABC competition). Therefore, there is 'knowing network' in which investigators just know each other or in the case of PIs, some colleagues are known to them. They may work in the same school or department, may have met in a conference, may have been working in the same field in different universities or even countries. Such a network should normally be expanding during an investigator's professional life as they have the time and opportunity to meet new people. At the same time, some parts of the network may get disconnected because of the changes in the subjects they are involved. Therefore, this kind of network is dynamic in both the members and the strength of the connections.

There exists a 'collaboration network' where people work or have worked with each other in some project. A PI chooses people from his/her knowing network and talks to them. Some people have the interest and time to become involved in the project team and they start a collaboration. New members may join through collaborators as members of their network; i.e., there may be members in the team that do not know the PI but get involved through collaborators. In the course of the investigation, new contacts are made and the team members get to know each other better, making some connections stronger for the future collaborations. One of the outcomes of collaborations is what is called social capital; investigators get a chance to know each other and new ideas are born in meetings, where potential collaborations and solutions are found. Therefore, collaborations both expand networks and social capital. Adding them to the model would allow for greater heterogeneity of performance.

The inclusion of collaboration would add lots of complexities to our model. In the first instance, it would be necessary to develop a decision rule to govern who would be picked as PIs who then would start gathering a team. The team size is another factor which should be decided and depends both on PI's characteristics and experience and the grant size. Decision making would be complicated as panels would now have information both on the PI and other members of the team; the weights for the PI and the members could be assigned but are likely to vary in practice. Different fields will assign creator rights differently. In science fields, PIs of labs often get credit on each paper resulting from their lab. Other disciplines are less clear about claims of co-production. Different rules could be used to decide how research results wold be assigned to team members (either equally or differently). Some collaborators also will likely be students who get training and experience and may become new collaboration candidates in future. So the community will change as time advances.

Some of the differences the inclusion of collaboration could bring about are as follows. There would be a dynamic network of investigators which could be studied to see how it changes during the simulation horizon. The changes will be an outcome of the model and can be studied along with the efficiency and equity aspects. The training of students is also another feature which can be studied along with the increment in the experience and knowledge of the investigation team. These can be studied under social capital which is a different output of the system than just papers and patents. However, the comparison between the paper/patent output and social capital (network expansion and training) would become a new problem.

In conclusion, it should be noted that any model is supposed to serve a purpose and that a compromise must be made between the complexity of a model and its applicability. Thus, there are occasions when a simple model provides enough insights into reality and, because of the greater costs involved, a more complex model is not used. It means that despite the fact that the above features can be included to make the model more realistic and complex, that may not be justifiable on economic grounds.

BIBLIOGRAPHY

Ahrweiler, P., A. Pyka, and N. Gilbert, (2004). "Simulating Knowledge Dynamics in Innovation Networks (SKIN)." In: Leombruni, R. and M. Richiardi (Eds.) *Industry and Labor Dynamics: The Agent-based Computational Economics Approach*. Proceedings of the Wild@Ace 2003 Workshop, Torino, Italy, 3-4 October 2003. Singapore: World Scientific Publishing.

Ahrweiler, P., A. Pyka, and N. Gilbert, (2011a). "A New Model for University-Industry Links in Knowledge-Based Economies." *Journal of Product Innovation Management*, 28(2): 218-235.

Ahrweiler, P., N. Gilbert, and A. Pyka, (2011b). "Agency and Structure: A Social Simulation of Knowledge-Intensive Industries." *Computational and Mathematical Organization Theory*, 17(1): 59-76.

Arvanitis, R., (2003) (Ed.) *Science and Technology Policy*. Vol 1. Encyclopedia of Life Systems. UNESCO Publishing/Eolss Publishers.

Auranen, O. and M. Nieminen, (2010). "University Research Funding and Publication Performance—An International Comparison." *Research Policy*, 39(6): 822-834.

Axelrod, R. and W. D. Hamilton, (1981). "The Evolution of Cooperation." *Science*, 211(4489): 1390-1396.

Ballot, G. and E. Taymaz, (1997). "The Dynamics of Firms in a Micro-to-Macro Model: The Role of Training, Learning and Innovation." *Journal of Evolutionary Economics*, 7(4): 435-457.

Ballot, G. and E. Taymaz, (1999). "Technological Change, Learning and Macroeconomic Coordination: An Evolutionary Model." *Journal of Artificial Societies and Social Simulation*, 2(2).3.

Bankes, S., R. Lempert, and S. Popper, (2002). "Making Computational Social Science Effective: Epistemology, Methodology, and Technology." *Social Science Computer Review*, 20(4): 377-388.

Boland, W. P., P. W. B. Phillips, C. D. Ryan, and S. McPhee-Knowles, (2012). "Collaboration and the Generation of New Knowledge in Networked Innovation Systems: A Bibliometric Analysis." *Procedia-Social and Behavioral Sciences*, 52: 15-24.

Bozeman, B. and M. Gaughan, (2007). "Impacts of Grants and Contracts on Academic Researchers' Interactions with Industry." *Research Policy*, 36(5): 694-707.

Cantner, U. and A. Pyka, (1998). "Technological Evolution —An Analysis within the Knowledge-Based Approach." *Structural Change and Economic Dynamics*, 9(1): 85-107.

Conte, R., N. Gilbert, G. Bonelli, C. Cioffi-Revilla, G. Deffuant, J. Kertesz, V. Loreto, S. Moat, J. P. Nadal, A. Sanchez, and A. Nowak, (2012). "Manifesto of Computational Social Science." *European Physical Journal-Special Topics*, 214(1): 325-346.

CRESS (Centre for Research in Social Simulation). SEIN page accessible at: cress.soc.surrey.ac.uk/skin/research/projects/sein. Accessed at 9th September 2015.

Dawid, H., (2006). "Agent-Based Models of Innovation and Technological Change." In: Tesfatsion, L. and K. L. Judd (Eds.) Handbook of Computational Economics: Agent-Based Computational Economics. (Vol. 2. Pp. 1235-1272). Chicago: Elsevier.

Dixon, P. M., J. Weiner, T. Mitchell-Olds, and R. Woodley, (1987). "Bootstrapping the Gini Coefficient of Inequality." *Ecology*, 68(5): 1548-1551.

Doern, G. B., D. Castle, and P. W. B. Phillips, (2016). *Canadian Science, Technology, and Innovation Policy: The Innovation Economy and Society Nexus*. Montreal: McGill-Queen's University Press.

Genome Canada, (2010). *Perspectives on Progress; Annual Report 2009-2010*. Ottawa: Genome Canada.

Genome Canada, (2016). *Corporate Plan 2016-2017*. Ottawa: Genome Canada.

Genome Canada, (2017). *Annual Report 2016-2017*. Ottawa: Genome Canada.

Genomics R&D Initiative, (2007). *Annual Performance Report 2006-2007*. Ottawa: Genome Canada.

Genomics R&D Initiative, (2014). *Annual Performance Report 2013-2014*. Ottawa: Genome Canada.

Gilbert, N., (1997): A Simulation of the Structure of Academic Science. *Sociological Research Online*, 2(2).3.

Gilbert, G. N., (2008). *Agent-Based Models*. London: Sage Publications.

Gilbert, N., A. Pyka, and P. Ahrweiler, (2001). "Innovation Networks –A Simulation Approach." *Journal of Artificial Societies and Social Simulation*, 4(3).8.

Gilbert, N., P. Ahrweiler, and A. Pyka, (2007). "Learning in Innovation Networks: Some Simulation Experiments." *Physica A: Statistical Mechanics and its Applications*, 378(1): 100-109.

Gilbert, N. and K. Troitzsch, (2005). *Simulation for the Social Scientist*. Berkshire, UK: Open University Press.

Grimm, V. and S. F. Railsback, (2005). *Individual-Based Modeling and Ecology*. Princeton: Princeton University Press.

Grimm, V., U. Berger, F. Bastiansen, S. Eliassen, V. Ginot, J. Giske, J. Goss-Custard, T. Grand, S. K. Heinz, G. Huse, and A. Huth, et al., (2006). "A Standard Protocol for Describing Individual-based and Agent-based Models." *Ecological Modelling*, 198(1): 115-126.

Grimm, V., U. Berger, D. L. DeAngelis, J. G. Polhill, J. Giske, and S. F. Railsback, (2010). "The ODD Protocol: A Review and First Update." *Ecological Modelling*, 221(23): 2760-2768.

Hakim, C., (2000). *Research Design: Successful Designs for Social Economics Research*. New York: Routledge.

Huffman, W. E., and R. E. Just, (1994). "Funding, Structure, and Management of Public Agricultural Research in the United States." *American Journal of Agricultural Economics*, 76(4): 744-759.

Heath, B., R. Hill, and F. Ciarallo, (2009). "A Survey of Agent-Based Modeling Practices (January 1998 to July 2008)." *Journal of Artificial Societies and Social Simulation*, 12(4)9.

Hinterberger, A., (2010). "The Genomics of Difference and the Politics of Race in Canada." In I. Whitmarsh and D. S. Jones (Eds): *What's the Use of Race? Modern Governance and the Biology of Difference*. Cambridge: MIT University Press.

ISED (Innovation, Science and Economic Development) Canada, (2015). *Evaluation of Industry Canada's Contribution To Genome Canada*. Ottawa: ISED.

Katz, J. S., (2006). "Indicators for Complex Innovation Systems." *Research Policy*, 35(7): 893-909.

McPhee-Knowles, S., (2014). The Complex Problem of Food Safety: Applying Agent-Based Modeling to the Policy Process. PhD. Dissertation, Johnson-Shoyama Graduate School of Public Policy, University of Saskatchwan, Saskatoon, Canada.

Merton, R. K., (1968). "The Matthew Effect in Science." *Science*, 159(3810): 56–63.

Merton, R. K., (1988). "The Matthew Effect in Science, II: Cumulative Advantage and the Symbolism of Intellectual Property." *Isis*, 79(4): 606-623.

OECD, (2012). *OECD Science, Technology and Industry Outlook 2012*. OECD publishing.

OECD, (2014). *OECD Science, Technology and Industry Outlook 2014*. OECD publishing.

Phillips, P. W. B. and E. Warren, (2010). "Managing Large-Scale Science Projects: The Genome Canada Experience." Working paper, www.VALGEN.ca.

Pyka, A., (2002). "Innovation Networks in Economics: From the Incentive-Based to the Knowledge-Based Approaches." *European Journal of Innovation Management*, 5(3): 152-163.

Pyka, A., N. Gilbert, and P. Ahrweiler, (2007). "Simulating Knowledge-Generation and Distribution Processes in Innovation Collaborations and Networks." *Cybernetics and Systems: An International Journal*, 38(7): 667-693.

Railsback, S. F. and V. Grimm, (2012). *Agent-Based and Individual-Based Modeling: A Practical Introduction*. Princeton: Princeton University Press.

Ryan, C. D., M. St. Louis, and P. W. B. Phillips, (2014). "Incorporating Network Analysis into Evaluation of 'Big Science' Projects: An Assessment Of The Canadian Light Source Synchrotron." *International Journal of Innovation Management*, 18 (4): 1450037.

Schelling, T. C., (1971). "Dynamic Models of Segregation." *Journal of Mathematical Sociology*, 1(2): 143-186.

Sen, A., (1973). *On Economic Inequality*. New York: Oxford University Press.

Sharma, P., (2012). Social Capital in Large-Scale Projects and Its Impact on Innovation: Social Network Analysis of Genome Canada (2000-2009). Master's Thesis, Johnson-Shoyama Graduate School of Public Policy, University of Saskatchwan, Saskatoon, Canada.

Statistics Canada (2016). Table 477-0017 - Number of Full-Time Teaching Staff at Canadian Universities, by Rank, Sex, Canada and Provinces, Annual (number). CANSIM.

STIC (Science, Technology and Innovation Council), (2015). State of the Nation 2014; Canada's Science, Technology and Innovation System: Canada's Innovation Challenges and Opportunities. Ottawa: Science, Technology and Innovation Council.

Teitelbaum, D. and H. Dowlatabadi, (2000). "A Computational Model of Technological Innovation at the Firm Level." *Computational & Mathematical Organization Theory*, 6(3): 227-247.

Triulzi, G. and A. Pyka, (2011). "Learning-by-Modeling: Insights from an Agent-Based Model of University–Industry Relationships." *Cybernetics and Systems*,42(7): 484-501.

Urban, J. B., N. D. Osgood, and P. L. Mabry, (2011). "Developmental Systems Science: Exploring the Application of Systems Science Methods to Developmental Science Questions." *Research in Human Development*, 8(1): 1-25.

Zhang, L. C., (2013). Policy Evaluation: A Case Study of Genome Canada Programming 2000-2011. Master's Thesis, Johnson-Shoyama Graduate School of Public Policy, University of Saskatchwan, Saskatoon, Canada.

Watts, C. and N. Gilbert, (2014). "Simulating Innovation: Computer-Based Tools for Rethinking Innovation." Cheltenham (UK): Edward Elgar Publishing.

Wikipedia (2016). "Excellence in Research for Australia." Accessed on December 24, 2016.