The Reputational Economics of Open Inputs Science

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

Duncan Halsall Law

BA (Hons) Cambridge, Peterhouse

School of Economics Finance and Marketing

College of Business

RMIT University

December 2019

Declaration

I certify that except where due acknowledgement has been made, the work is that of the author alone; the work has not been submitted previously, in whole or in part, to qualify for any other academic award; the content of the thesis is the result of work which has been carried out since the official commencement date of the approved research program; any editorial work, paid or unpaid, carried out by a third party is acknowledged; and, ethics procedures and guidelines have been followed.

I acknowledge the support I have received for my research through the provision of an Australian Government Research Training Program Scholarship.

Duncan Law

08 December 2019

Acknowledgements

I would like to thank my supervisors, Jason Potts, Prateek Goorha, and Chris Berg, for their guidance, encouragement, and for the freedom they gave me to explore the intellectual terrain of the thesis. The economics reading group that Jason and Prateek led through the early years of the thesis research was particularly important in expanding the intellectual scope of my research, and for deepening its engagement with new institutional economics.

Other members the RMIT innovation commons team - in particular Darcy Allen and Trent MacDonald - provided valuable feedback in conversations about innovation economics, Coase, Ostrom, and many other topics. Stefanie Schurer gave me important early pointers around formal modelling approaches. My discussions of agent-based modelling with Liam Magee were important in selecting the modelling strategies I eventually adopted in the thesis. Conversations with Mike Beggs, Carl Dyke, and Nate Holdren have helped to develop my thinking about institutional political economy.

I acknowledge the support I have received for my research through the provision of an Australian Government Research Training Program Scholarship

I am grateful to my parents, John Law and Sheila Halsall, for their lifelong support - and to John for fostering an abiding interest in science studies, as well as for patiently tolerating my regression to Mertonian analysis. Finally, Nicole Pepperell, in innumerable conversations and several co-authored papers, has provided multiple rounds of supportive but critical feedback: our discussions of science, fallibilism, pragmatism, and institution design have shaped the argument of the thesis in deep ways.

Contents

Abstract **1**

Introduction **3**

I **12**

1 Defining open science **13**
1.1 Science and innovation 14
1.1.1 From basic research to economic growth: the linear model 15
1.1.2 Alternatives to the linear model 16
1.2 Four meanings of 'open science' 21
1.2.1 Open findings 22
1.2.2 Open access 24
1.2.3 Open inputs 27
1.2.4 Citizen science 29
1.3 The focus of the thesis 30

2 The economics of science **33**
2.1 Introduction 33
2.2 Early neoclassical economics of science 34
2.3 Knowledge and growth theory 39
2.4 The new economics of science 43
2.4.1 Robert K. Merton's sociology of science 44
2.4.2 Science as prisoner's dilemma 47
2.5 The economics of scientific knowledge 49
2.5.1 Truth as consensus: symmetry in the ESK 51

2.5.2 Evolutionary social epistemology 55
2.6 Agent-based modelling 57
2.6.1 Agent-based models of science 59
2.7 Summary and conclusion 64

3 Modelling institutions **66**
3.1 Introduction 66
3.2 Institutional economics 67
3.2.1 Old institutional economics 68
3.2.2 Coase 69
3.2.3 New Institutional Economics 72
3.3 The analytic status of the 'rules of the game' 73
3.3.1 Elinor Ostrom's Institutional Analysis and Development approach 76
3.4 Institutions as games and networks 79
3.4.1 Science as non-cooperative game 80
3.4.2 Science as epistemic network 83
3.4.3 Summary 87
3.5 Four concepts of modelling 88
3.6 Epistemology and science studies 92

II **94**

4 The rules of the scientific game **95**
4.1 Introduction 95
4.2 Science as a global institution 96
4.2.1 The republic of science 97
4.3 Commons within the republic of science 101
4.3.1 Innovation commons 105
4.3.2 Scientific innovation commons 107
4.4 The internal structure of Mertonian science 111
4.4.1 Three categories of scientific commons 113

5 Modelling the republic of science **117**
5.1 Introduction 117
5.2 Cooperative outcomes in the n-player prisoner's dilemma . . . 118
5.2.1 The iterated n-player prisoner's dilemma 119

5.3 Conditionally cooperative strategies and the Ostromian commons . . . 122
5.4 Evolutionary game theory . . . 124
5.5 Description of the model . . . 125
5.6 Findings . . . 128
5.7 Discussion . . . 131

6 The opinion dynamics of Mertonian science 133
6.1 Introduction . . . 133
6.2 David's invisible college voter model . . . 134
6.3 A veritistic approach to modelling invisible colleges . . . 137
6.4 The simple dynamics of the invisible college model . . . 139
6.5 The interaction of the college with the broader public sphere . 143
6.6 Limitations of the basic invisible college model . . . 148
6.7 The CODA model . . . 150
6.8 The opinion dynamics of the simple CODA model . . . 152
6.9 Conclusion . . . 158

III 161

7 The rules of the open inputs game 162
7.1 Introduction . . . 162
7.2 Debating open inputs science . . . 163
7.2.1 Open inputs and scientific efficiency . . . 163
7.2.2 Open inputs and scientific reliability . . . 164
7.2.3 Evaluating open inputs science . . . 168
7.3 Obstacles to open inputs science . . . 170
7.4 Reputational constraints on intermediate input sharing . . . 172
7.5 Open inputs policies . . . 174
7.5.1 The problem of incentive compatibility . . . 179

8 Modelling open inputs science 182
8.1 Introduction . . . 182
8.2 The prisoner's dilemma of open inputs science . . . 183
8.2.1 Extending the evolutionary model . . . 184
8.3 The opinion dynamics of open inputs science . . . 187
8.3.1 Reviewing the CODA model . . . 189

8.3.2 Open inputs in the CODA model 192
8.3.3 Discussion 195
8.4 Summary 197

9 Conclusion **198**
9.1 Recapitulation of the argument 198
9.2 Implications of the argument 202
9.3 Limitations and opportunities 204

Bibliography **207**

List of Figures

3.1 The Institutional Analysis and Development framework, as depicted in Strandburg et al. (2017), p. 15, after Ostrom (2005). Licensed under a Creative Commons Open Access licence CC-BY-NC-ND 4.0. To view a copy of this license, visit https://creativecommons.org/licenses/by-nc-nd/4.0. . . . 77

4.1 The Institutional Analysis and Development framework for studying common pool resource governance communities, as depicted in Strandburg et al. (2017), p. 15, after Ostrom (2005). Licensed under a Creative Commons Open Access licence CC-BY-NC-ND 4.0. To view a copy of this license, visit https://creativecommons.org/licenses/by-nc-nd/4.0. . . . 103
4.2 The modified Institutional Analysis and Development framework, for studying knowledge commons governance communities, as depicted in Strandburg et al. (2017), p. 15. Licensed under a Creative Commons Open Access licence CC-BY-NC-ND 4.0. To view a copy of this license, visit https://creativecommons.org/licenses/by-nc-nd/4.0. . . . 104

5.1 Examples of n-player prisoner's dilemmas 121
5.2 The user interface of the NetLogo model, with an illustrative graph of 90 agents. 127
5.3 Representative simulations. The green lines indicate the number of cooperative agents; the red lines the number of non-cooperative agents. 129
5.4 Average number of cooperators after 300 rounds of play 130

6.1 David's representation of the simple invisible college graph. . . 134
6.2 The basic invisible college network used in this chapter. 135

6.3 An example of the basic invisible college model in NetLogo. . . 140
6.4 Number of agents believing P after 1000 rounds across 1000 simulations: the simple voter model. 141
6.5 Number of agents believing P after 1000 rounds of play across 1000 simulations: with observation of nature. 142
6.6 Scientists and citizens sharing a broader public sphere. 144
6.7 Percentage of agents believing P after 1000 rounds, across 1000 simulations: scientists within the public sphere 146
6.8 Percentage of agents believing P after 1000 rounds, across 1000 simulations: citizens within the public sphere 147
6.9 Evolution of agents' opinions 152
6.10 Basic CODA model: distribution of agents' likelihood estimates of P, after 1000 rounds of play 153
6.11 CODA model with observation of nature: Percentage of agents believing P after 1000 rounds of play, for a range of different probabilities of agents carrying out experiments 154
6.12 'Pessimistic' CODA model: Percentage of agents believing P after 1000 rounds of play, for a range of different experiment probabilities, given an initial 'bad' consensus 155
6.13 CODA model with retirement: Percentage of agents believing P after 1000 rounds of play, for a range of different experiment probabilities and retirement rates 157

8.1 Shift from closed to open input sharing, as the relative costs of closed and open cooperation change 188
8.2 CODA model: As the rate of experimentation increases, the percentage of agents believing P changes. 190
8.3 CODA model with retirement: Different retirement rates result in different 'tipping points' at which agents shift to believing P. 191
8.4 CODA model with open inputs science: Each line represents the mean percentage of agents believing P, across a range of different rates of experimentation, for a given rate of open input sharing. This chart represents a rate of agent retirement of 2% 193
8.5 CODA model with open inputs science: Agent retirement rate of 5% 194

Abstract

This thesis is concerned with the problem of how best to organise the institutions of scientific research. In particular, it is focused on the problems raised by contemporary debates over 'open inputs science': the broad sharing of the intermediate inputs of scientific production on the open web. The problem motivating the thesis can be stated as follows: what institutional structures can incentivise *efficient* scientific production, via the broad cooperative sharing of *intermediate inputs*, subject to the constraint that the *credibility* of scientific output must be maintained or improved?

The thesis draws on the resources of new institutional economics, and uses the tools of agent-based modelling, to develop a series of illustrative models of scientific institutions. I use these models to make a multi-stage argument about the structure of science as a global institution, and the implications of this institutional structure for 'open inputs science'.

At the most abstract level, the thesis analyses two distinct categories of scientific institution. First, the thesis analyses traditional 'Mertonian' science, in which scientists compete within a reputational economy for priority of publication. Within this reputational economy, the publication of scientific findings is incentivised, but the broad sharing of the intermediate inputs of scientific production is *dis*incentivised.

Second, the thesis analyses the institutional structures and proposals associated with 'open inputs science'. These 'open inputs' proposals seek to incentivise *both* open publication of research findings *and* the widespread sharing of intermediate inputs like data, code, research protocols, etc. Open inputs advocates and practitioners therefore face the challenge of transforming the incentive structures of traditional, Mertonian science into an institutional system that can encourage broader open input sharing.

The thesis analyses these categories of institutional structure along two dimensions. First, I am interested in *cooperation* within scientific production.

I build on Dasgupta and David's analysis of science as a prisoner's dilemma, to explore the institutional scenarios in which different forms of cooperative behaviour are likely to spread across a community of practising scientists. I argue that a scientific community structure in which many small-scale sub-communities are *weakly* bound together within a larger network can foster the spread of cooperative practices.

Second, I am interested in the epistemic outcomes associated with different scientific institutions. I draw on the social epistemology literature, and use agent-based opinion dynamics modelling, to study the impact of specific institutional changes on scientific epistemic outcomes. I argue that there is good reason to believe that open inputs science has a significant positive epistemic impact on scientific consensus-formation.

Finally, alongside these formal modelling dimensions of the thesis, I also draw on and contribute to the qualitative literature on science as commons. In particular, I apply the concept of the 'innovation commons' to the study of scientific institutions, and I argue that policymakers interested in achieving open inputs science might be well served by a greater focus on the small-scale innovation commons as a locus of institutional transformation.

Taken all together, these elements of the thesis provide a complex synthetic account of the interrelation between different subcomponents of science as a global institution, and the implications of this institutional structure for the project of 'open inputs science'.

Introduction

This thesis is concerned with the problem of how best to organise the institutions of scientific research. In particular, it is focused on the problems raised by contemporary debates over 'open science'. The thesis takes a comparative institutional approach to the analysis of scientific production. I argue that such an approach can illuminate the study of scientific institutions *both* by allowing us to see existing and traditional scientific institutions in a new light, informed by present concerns, *and* by expanding the resources we have available for the contrastive analysis and evaluation of contemporary institutional proposals. More concretely, the problem motivating the thesis can be stated as follows: what institutional structures can incentivise *efficient* scientific production, via the broad cooperative sharing of *intermediate inputs*, subject to the constraint that the *credibility* of scientific output must be maintained or improved?

In the body of the thesis I ask and answer this question in relation to a set of specific scientific institutions. In particular, I make use of theoretical resources drawn from the 'new economics of science' and the 'economics of scientific knowledge', as well as the analytic toolkit of agent-based modelling, to address current policy debates over the potential strengths and weaknesses of recent 'open science' institutional proposals.

The thesis consists of three parts. In Part I, I provide the necessary background, context, and theoretical resources. In Part II, I develop a series of models of traditional, 'Mertonian' scientific institutions. Then in Part III, I make use of those models to analyse and evaluate a set of contemporary institutional proposals typically grouped under the heading 'open science'.

In this introduction I provide an overview of the argument of the thesis as a whole. The goal in this introduction is not to make the case for that argument - rather, my goal is to give a sufficiently elaborated sense of the overall structure of the thesis, to make clear the motivation for the compo-

nents of the argument that I will assemble in later chapters.

The thesis is motivated by the problems of scientific institution design. Specifically, it is motivated by two problems that all scientific institutions must aim to address. On the one hand, scientific institutions aspire to *efficiency* in the research process: reducing the unnecessary duplication of labour and investment, minimising the extent to which resources are deployed on fruitless projects, etc. On the other hand, scientific institutions aspire to *credibility* of output. A characteristic feature of science - one of the features that marks scientific output as distinctively scientific - is its ability to assert a particular kind of empirical and rational warrant for the credibility of its claims. In terms of the output of the community as a whole, science that is not credible is not science. One of the fundamental problems of scientific institution building, therefore, is the goal of achieving both high efficiency of process, and high credibility of output - and of these two desiderata, credibility of output is the most essential.

These problems of scientific institution design have been highlighted in recent decades by a set of debates over the best way to organise scientific production: debates over the question of 'open science'. In chapter one of the thesis, I place those debates in context, by outlining a set of rival accounts of the role of science in innovation. I give an account of the 'linear model' of the relationship between science and production, and describe a series of critiques of that model, and alternative proposals. I argue that there is both a broad consensus about the importance of science to economic growth, and a *lack* of consensus about exactly what role scientific institutions play, or should play, in the knowledge economy.

I then focus on a specific set of proposals for transforming scientific institutions: those typically grouped under the umbrella term 'open science'. I argue that this umbrella term captures *four* distinct classes of institutional proposal, and I provide an overview of specific proposals within that schematisation. This allows me to specify the sense of 'open science' that will be the primary focus of the thesis. I call this *open inputs* - the broad public sharing, with low barriers to access, of the intermediate inputs of scientific production.

In chapter two of the thesis, I then pull back to review the theoretical literature in the economics of science. I summarise the mid-twentieth century neoclassical consensus, and discuss distinct theoretical traditions that have challenged and built upon that framework: endogenous growth theory; the

'new economics of science'; the 'economics of scientific knowledge'; and agent-based modelling of scientific research dynamics.

In particular, the 'new economics of science', exemplified by the work of Partha Dasgupta and Paul David, looks at a central problem of inefficiency in scientific production. For Dasgupta and David, drawing on the sociological work of Robert K. Merton, scientists are primarily motivated by *reputation*, and they achieve reputation by means of *priority of publication*. This institutional structure incentivises open publication - 'open findings' in the sense defined in chapter one - but it disincentivises the sharing of intermediate inputs among scientific researchers. Although the community as a whole would benefit from collective input pooling - this is the most efficient collective practice - every individual scientist benefits from hoarding inputs. Mertonian science can therefore be modelled as an n-player prisoner's dilemma. This provides a formal articulation of one dimension of the central problem that motivates the thesis: how can scientific institutions generate cooperative outcomes in the iterated n-player prisoner's dilemma of intermediate input sharing?

The 'economics of scientific knowledge' then provides the resources to elaborate the second dimension of the problem: the question of credibility. The economics of scientific knowledge uses the formal resources of economics to model the impact of different scientific institutional structures on the *epistemic outcomes* of scientific production. I argue that Alvin Goldman's veritistic social epistemology provides a metatheoretical framework for the comparative epistemic evaluation of different scientific institutional structures. I further argue that, within this framework, the toolkit of opinion dynamics modelling allows us to model a range of different scientific institutional structures. Finally, I argue that the broader resources of agent-based simulation allow us to explore these theoretical perspectives within a dynamic modelling framework well-suited to the study of complex heterogeneous communities like those of scientific researchers. The joint use of these various resources therefore provides the core theoretical toolkit necessary to address the problem of the thesis.

Having outlined these theoretical resources, in chapter three I pan back again, to discuss some of the challenges of modelling institutions. I give an overview of the tradition of institutional economics, with a particular emphasis on the work of Ronald Coase as a bridge between the 'old' and 'new' institutional economics. I schematise new institutional economics into two subtraditions, focused on transaction costs, on the one hand, and the 'rules of

the game' that structure social and economic practice, on the other. I suggest that Elinor Ostrom provides a particularly valuable toolkit for the qualitative analysis of how historically and socially contingent institutions shape these 'rules of the game'. I then discuss some of the more formal resources available for institutional economics, outlining a range of approaches to game-theoretic and opinion dynamics modelling.

Finally, I address a set of methodological and epistemological problems associated with the modelling strategies adopted in the thesis. I discuss the problem of realism in economic modelling, and I suggest that the models presented in this thesis can best be understood as 'caricatures', in the sense defined by Gibbard and Varian: simplified representations of one dimension of social reality, which gain their methodological warrant from the explanatory clarity this simplicity provides. Then I discuss some of the epistemological problems associated with the social-scientific study of science. In the debates over relativism, constructivism and realism in science studies, this thesis adopts a stance of 'weak methodological realism'. This epistemological approach aims to enable the comparative epistemic evaluation of scientific institutions, while largely bracketing the metaphysical or metatheoretical questions associated with the problem of scientific objectivity.

These metatheoretical and epistemological considerations conclude my survey, in Part I, of the theoretical resources required for the substantive project of the thesis. In Part II, I then begin to carry that project out, by developing a set of models of traditional, 'Mertonian' scientific institutions. Part II as a whole uses these models to analyse the approaches taken by traditional, Mertonian science to resolve the dual institution-design problems of intermediate input sharing and credibility of output.

The argument of Part II has three stages. In chapter four, I discuss the 'rules of the game' of traditional, 'Mertonian' science. This discussion of Mertonian science takes the work of Dasgupta and David as its starting point, and extends the approach in three ways. *First*, I connect Dasgupta and David's account to the analysis of commons governance institutions developed by Elinor Ostrom - and in particular the application of Ostrom's approach to scientific institutions that has recently been developed and pursued by the Workshop on Governing Knowledge Commons. *Second*, I connect this concept of science as commons to Michael Polanyi's idea of the Republic of Science. *Third* and finally, I discuss Jason Potts' concept of an *innovation commons*, and argue that this concept illuminates one of the key components of the scientific innovation ecosystem: the 'invisible college' as Ostromian in-

novation commons. All of these resources taken together, I argue, enable us to embed the basic Dasgupta and David model of Mertonian science within a more complex account of science as a large-scale institutional system.

This picture of science as a complex system in turn allows us to do two things. First, it lets us distinguish between three distinct categories of scientific commons: open knowledge commons, in which information is shared via the public sphere; closed knowledge commons, in which information is shared exclusively within a small-scale community or club; and Ostromian commons, as 'rules of the game' that structure the communities producing and maintaining knowledge commons. Second, it gives us a qualitative picture of the 'rules of the scientific game' that can be used as the basis for formal, computational models of the dynamics of science as system. In the remainder of Part II, I use this qualitative picture of science as an institution to develop a set of agent-based models of scientific dynamics.

In chapter five, I develop an agent-based model of intermediate input-sharing within Polanyi's 'republic of science'. Polanyi's idea of the republic of science suggests that the global scientific community is best understood, not just as a *single* community, but as a vast network composed of many small, partially overlapping communities. In chapter five I use agent-based modelling to construct a graph-theoretic environment that resembles Polanyi's republic of science. I then run simulations in which many n-player prisoner's dilemmas are played by subcommunities within this environment, to model the problem of cooperative input sharing within many small-scale 'invisible colleges'. By altering the parameters of the model, I can assess which community structures most reliably produce general cooperative outcomes - I find that *strong* ties *within* subcommunities, together with *weak* ties *between* subcommunities, reliably result in the spread of cooperative strategies across the community as a whole, via a form of 'group selection'. This finding, I argue, shows that Polanyi's republic of science can enable the spread of cooperative practices within 'invisible colleges', as described in chapter four, even across a global scientific community of arbitrarily large size.

Chapter six then turns to the second evaluative dimension of the thesis: the problem of credibility. I outline a set of opinion dynamics models of scientific consensus formation, drawing on the work of Paul David and A.C.R. Martins. I use these approaches to model the epistemic behaviour of an ideal-typical 'republic of science', under a range of different behavioural scenarios.

I first consider the behaviour of a scientific community in which scientists are motivated 'purely' by reputational outcomes. Then I consider a com-

munity in which scientists attempt to form true beliefs about the natural world using rational Bayesian update principles. Both of these scenarios, I conclude, can potentially result in 'herd behaviour', and thus in negative epistemic outcomes. However, rational, Bayesian scientists may be *more* vulnerable to herd behaviour, than are reputationally-motivated scientists with simpler 'rule of thumb' update rules. This apparently counter-intuitive result is compatible with the related work of Philip Kitcher on scientific epistemic incentive structures and suggests one functionalist warrant for the 'reputational' incentive structure of Mertonian science, discussed in Part I.

I next consider the problem of the relationship between the community of scientists and the broader community of non-scientist citizens. Here I extend David's model of scientific 'invisible colleges' to incorporate the relationship between scientific and non-scientific epistemic agents. This model also implies a seemingly counter-intuitive result: under certain conditions, epistemic outcomes may be superior among citizens than among scientists, provided those citizens' opinions are informed by scientific output. This result - though dependent on the parameters of the model - is made possible by the *fallibilism* intrinsic to reliable scientific research: because science relies on a process of rational, empirical evaluation of competing hypotheses, a scientific community must be capable in principle of 'disrupting' its own consensus, and therefore will not reliably result in uniform true belief among its members.

In sum, the chapters of Part II develop a series of models of traditional, 'Mertonian' science. Taken together, these models provide an account of how the scientific 'rules of the game' are established in institutional practice; how cooperative behaviour can be achieved within these rules; and how these rules influence the epistemic outcomes of the system as a whole.

In Part III, I turn once again to the ways in which this institutional structure is being challenged by contemporary open science proposals. Advocates of 'open inputs science' argue that cooperative input sharing *within* 'invisible colleges' is insufficient: intermediate inputs should instead be shared with the *entire* scientific community, and with people outside that community, via their free publication on the open web. For open input advocates, the institutional problem is how to encourage this open input sharing, given the incentive structure of existing Mertonian science.

In chapter seven I outline the institutional changes and proposals associated with the shift to open inputs science. As in chapter three, I argue that these shifts can be understood in terms of transaction costs, on the one

hand, and 'rules of the game', on the other. The development of new technologies, particularly associated with the internet, have dramatically reduced the transaction costs associated with many forms of input sharing. However, the specific changes to the scientific 'rules of the game' occasioned by those shifts in transaction costs remain a matter of ongoing debate. I use the example of the recent 'replicability crisis' in psychology as a window onto some of these debates.

I then use the models of Mertonian science developed in Part II, to make a series of original contributions to these debates, by extending each of these models to incorporate the possibility of open input sharing.

First, I extend the evolutionary game-theoretic model of chapter five to incorporate a new game strategy: open input sharing. I discuss the scenarios under which this strategy outcompetes the traditional strategy of closed inputs science, and conclude that relatively modest shifts in institutional norms can make open input sharing a viable strategy. Significantly, this model suggests that one key locus of incentive transformation is the small-scale 'innovation commons' - a finding that has potential consequences for science policy.

I then extend the opinion dynamics models of chapter six, to incorporate the possibility of open inputs sharing. This original model suggests that open input sharing can indeed produce a positive step change in epistemic outcomes, but that the frequency and reliability of scientists' observations of nature remain the major determinants of overall epistemic outcomes.

In the conclusion of the thesis, I draw out some of the implications of this comparative institutional analysis, I discuss how the analysis carried out in the thesis might inform ongoing debates around scientific institution design, and I sketch directions for future research building on this approach.

This survey of the structure of the thesis is not intended to substitute for the detailed arguments and models developed in subsequent chapters. Its purpose, rather, is to provide a context for the pieces of the argument I will begin to assemble in Part I. Focusing, in this preliminary way, on the twin problems of scientific efficiency and scientific credibility is intended to make clear the overall trajectory of the thesis - and the relevance of its components to a single overarching problem about the possibilities for scientific institutional design and change.

Some earlier versions of portions of the argument below have appeared in the following publications and presentations:

> Law, Duncan (2014). Scientific boundary-work and the contested status of open science. Paper presented at The Australian Sociological Association Conference 2014: Challenging Identities, Institutions and Communities. University of South Australia, Adelaide, 27 November 2014.
>
> Law, Duncan and Nicole Pepperell (2014). "Knowledge and social practice: Robert Brandom and the strong programme". In: *Refereed Proceedings of the Australian Sociological Association Conference 2014.* TASA 2014: Challenging Identities, Institutions and Communities. Adelaide: The Australian Sociological Association (TASA), pp. 1-19.
>
> Pepperell, Nicole and Duncan Law (2015). The internet imaginary: Between technology and technique. In: *M/C Journal*, 18.2. URL: `http://journal.media-culture.org.au/index.php/mcjournal/article/view/957`.
>
> Law, Duncan (2016). Theorising internal institutional diversity: Elinor Ostrom on rules and social conflict. In: *Refereed Proceedings of TASA 2016 Conference.* TASA 2016: Cities and Successful Societies. Melbourne: The Australian Sociological Association (TASA), pp. 206-212.
>
> Law, Duncan and Nicole Pepperell (2017). Sociology and the mirror of nature: Robert Brandom and the strong programme. In: *Journal of Sociology* 53.1, pp. 245-263.
>
> Law, Duncan and Nicole Pepperell (2018). Oppression in the commons: Cautionary notes on Elinor Ostrom's concept of self-governance. In: *Refereed Proceedings of TASA 2018 Conference.* TASA 2018: Precarity, Rights and Resistance. Melbourne: The Australian Sociological Association (TASA), pp. 7-12.

The NetLogo code associated with this thesis is available for download at the NetLogo modelling commons: `http://modelingcommons.org/account/models/3578`

The model simulations data used to generate the figures in the thesis is available for download at: `https://tinyurl.com/ydjl783e`

Part I

Chapter 1

Defining open science

In Part I of the thesis, I lay the groundwork for the later analysis by providing the necessary background, context and theoretical resources. I begin that task, in this first chapter, in two ways. First I motivate the focus of the thesis on scientific institutions by describing a set of debates over the role of science in the broader innovation process. In section 1.1, I argue that, while there is consensus in the contemporary scholarly literature on the importance of scientific research to innovation, there is a lack of consensus about exactly what role science plays in the knowledge economy. To do this, I summarise the mid-twentieth century 'linear model', a range of criticisms of that model, and a set of alternative proposals.

This lack of consensus is exacerbated by an ongoing set of debates over the appropriate institutional structure within which scientific research should be carried out. In section 1.2, I focus on the specific institutional proposals that will be the focus of the thesis - those often grouped under the umbrella term 'open science'. This term captures a very wide range of institutional proposals. I schematise these into four distinct classes of proposal, under the headings 'open findings', 'open access', 'open inputs', and 'citizen science'. The primary focus of the remainder of the thesis will be *open inputs*.

In chapter two, I then give an overview of the economics of science - with a particular focus on the 'new economics of science' and the 'economics of scientific knowledge. These two traditions, I argue, provide the theoretical resources necessary to formally model the scientific institutions and debates that I will be studying in later chapters. I survey recent literature on the agent-based modelling of science - the major modelling approach I will be drawing on in later chapters.

In chapter three I pull back again, to review the literature on modelling institutions. I look at the tradition of institutional economics, including the distinction between 'old' and 'new' institutional economics, and I outline a set of additional modelling resources for the comparative analysis of scientific institutions. Finally, I discuss some methodological and epistemological questions associated with these types of economic modelling.

Taken all together, the material discussed in the three chapters of Part I provides the resources required to carry out the comparative analysis of traditional and 'open inputs' scientific institutions, which forms the focus of the remainder of the thesis.

1.1 Science and innovation

Policy-makers value and economists study scientific research because of the impact of scientific findings on technological innovation, economic growth, and human welfare. In Richard Freeman's words:

> If you believe that long- term economic growth depends on the application of science and technology to the production of goods and services — and who in the knowledge economy does not? — then science lies at the heart of economic growth. (Freeman 2011: 85)

Growth in turn is a central preoccupation for economists and policy-makers because of its profound impact on human welfare. As Robert Lucas writes, discussing the potential impact of economic policy on the growth rate differential between national economies:

> The consequences for human welfare involved in questions like these are simply staggering: Once one starts to think about them, it is hard to think about anything else. (Lucas 1988: 5)

I discuss Lucas's contributions to growth theory in more detail in chapter two. In this section I have two narrower goals: first, to demonstrate the broad consensus that science is a crucial component of economic growth; second, to demonstrate the *lack* of consensus around exactly how to analyse the role of science in the broader economy. I do this by first discussing the linear model of the connection between science and innovation, which was dominant in the middle decades of the twentieth century, then summarising a range of more recent critiques of and departures from that model.

1.1.1 From basic research to economic growth: the linear model

In his 1945 text 'Science: The Endless Frontier' Vannevar Bush articulated an understanding of the role of science in the national economy that was influential among both policy-makers and economists in the middle decades of the twentieth century. In that document Bush draws a distinction between basic and applied research, and proposes that heavy investment in the former is a precondition for the ongoing success of the latter in the development of new products and production processes. For Bush, moreover, the government is the economic actor best placed to provide such investment in basic research. Bush therefore proposes the creation of a government agency - the U.S. National Science Foundation - to fund and coordinate basic research, modelled on the wartime Office of Scientific Research and Development.

For Bush, such coordinated government investment in basic research is a fundamental requirement for ongoing economic success, grounded in technological advance. In Bush's words:

> The scientist doing basic research may not be at all interested in the practical applications of his work, yet the further progress of industrial development would eventually stagnate if basic scientific research were long neglected... Basic research leads to new knowledge. It provides scientific capital. It creates the fund from which the practical applications of knowledge must be drawn. New products and new processes do not appear full-grown. They are founded on new principles and new conceptions, which in turn are painstakingly developed by research in the purest realms of science. (Bush 1945: 13-14)

Bush's text is a work of policy advocacy, not economic theory, but it is widely seen as articulating an understanding of the relationship between scientific research and economic growth that enjoyed broad consensus in the middle decades of the twentieth century: the so-called 'linear model'. In his intellectual history of the development of this model, Godin (2006) argues that "Bush borrowed his arguments directly from his predecessors, among them industrialists and the United States National Research Council" (640), and that the model which underpins Bush's report was further extended and refined in the years following the report's publication. In Godin's words, the full linear model, as it was articulated in the late 1940s and 1950s:

> postulates that innovation starts with basic research, then adds applied research and development, and ends with production and diffusion:
>
> > Basic research → Applied research → Development → (Production and) Diffusion (639)

The linear model thus draws a sharp distinction between a series of stages of the production process, and proposes a largely unidirectional causal chain between those stages.

This model was not just articulated, as it was by Bush, in macro terms, but also in meso and micro forms (e.g. Furnas 1948; Maclaurin 1953). Thus, while the linear model lacked a single canonical articulation, it provided a framework for understanding the role of science at all scales of the economy, shared by many economists, managers and policy-makers from the 1940s.[1]

From the 1980s onward, however, the linear model became the target of a range of critiques. As Balconi, Brusoni, and Orsenigo (2010) have argued, not all of these critiques were fair: many targeted a stereotype of the model that few of its advocates had ever held. Nevertheless, widespread scepticism towards the linear model led to a development of a range of newer approaches that continue to centrally inform debate. In the following subsection I give an overview of some of the most prominent criticisms of the linear model, and alternative accounts of the connection between science and innovation.

1.1.2 Alternatives to the linear model

The linear model discussed in the previous subsection had two components: first, a set of distinctions between different elements of the production process; second, the idea that these elements could be organised into a largely unidirectional causal chain. From the 1980s onward, critics of the linear model challenged both of these ideas.

[1] Edgerton (2004) controversially argues that the linear model, in the form just described, never existed - that it is merely "a foil for the more elaborated academic account, in short, a classic straw man" (31). Godin's (2006) intellectual history of the development and spread of the model provides a comprehensive rebuttal of that suggestion, as does Balconi et al.'s (2010) schematisation of different historical versions of the model, and defence of a 'weak' form of the model. See also Stevens (1941), Furnas (1948), Maclaurin (1953) and Ames (1961) for mid-twentieth century examples of versions of the model 'in the wild'.

One set of critics argued that the linear model's account of the distinction and relationship between basic and applied research fails to capture important elements of the actual research process. As many scholars have emphasised (e.g. Kline and others 1985; Gardner 1995; Krige 2006), the history of science contains many instances of basic insights arising from applied research, rather than vice versa. Kline, for example, discusses how trial and error experimentation often produces findings that precede analysis and explanation of the underlying mechanisms. In a recent history of the development of the aerofoil, the sociologist of science David Bloor discusses how the scientific task was to explain an already functional technology (Bloor 2011). In Clifford Geertz's words: "Science owes more to the steam engine than the steam engine owes to science; without the dyer's art there would be no chemistry; metallurgy is mining theorized" (Geertz 1980: 168).

Critics did not just challenge the linear model's account of the causal link between basic and applied research; they also argued that the basic/applied distinction itself is flawed. Thus Donald Stokes, in a 1994 paper and subsequent book, argues that the basic/applied distinction conflates two separate distinctions that are better analysed independently (Stokes 1994, 1997). For Stokes, the question of whether research aspires to fundamental understanding is orthogonal to the question of whether research is carried out with the goal of useful applications. Stokes therefore proposes a two-by-two matrix of research types, the quadrants of which he associates with paradigmatic historical researchers.

		Considerations of use?	
		No	Yes
Quest for fundamental understanding?	Yes	Pure basic research (Bohr)	Use-inspired basic research (Pasteur)
	No		Pure applied research (Edison)

For Stokes, 'Pasteur's Quadrant' - research that is oriented to both the resolution of immediate practical challenges and the expansion of fundamental scientific understanding - is a major class of research overlooked by many applications of the linear model.

These researchers challenged the linear model's claims about the relation

between basic and applied research. But researchers also challenged the linear model's broader picture of the the production process. In a series of works, von Hippel (Von Hippel 1986, 2001; Von Hippel and Von Krogh 2003; cf. Thomke and Von Hippel 2002) emphasises the central role played in the innovation process by economic actors outside the traditional sites of production - especially the importance of customer or user innovation. Von Hippel argues both that the importance of user innovation has long been underestimated, and that new technologies and economic structures - such as the open source software community - are 'democratising innovation'. This insight has prompted scholars to focus on the ways in which user innovation can be more explicitly incorporated into our understanding of the innovation process (Franke and Schreier 2002; Franke and Piller 2004).

Even for researchers who remain focused on traditional sites of production like the firm, however, the linear model's understanding of the role of research in product development is often seen as flawed. For Kline and Rosenberg, for example, "the central process of innovation is not science but design" (Kline and Rosenberg 1986: 286). Kline and Rosenberg do not contest the importance of science in innovation, but they argue that a more complex account of the relevant causal chains - and feedback loops - is necessary. They propose a 'linked-chain' or 'chain-linked' model with a five-step innovation process: market finding; invention and/or analytic design; detailed design and testing; redesign and production; and marketing and distribution. Each of these steps is moved through in turn, with the potential for feedback loops between them. Scientific research, in turn, can influence *each* of the five steps, with the *exception* of the first. In Kline and Rosenberg's words:

> the use of the accumulated knowledge called modern science is essential to modern innovation; it is a necessary and often crucial part of technical innovation, but it is not usually the initiating step. It is rather employed at all points along the central-chain-of-innovation, as needed. It is only when this knowledge fails, from all known sources, that we resort to the much more costly and time-consuming process of mission-oriented research to solve the problems of a specific development task. (291)

The preceding criticisms of the linear model are primarily focused on the micro or meso levels of analysis. Another group of scholars have challenged the linear model at a more macro level: For these researchers, the linear model

primarily failed because it did not address the complexity of the broader innovation ecosystem.

Thus the 'National Systems of Innovation' (or 'National Innovation Systems' - NIS) approach, developed from the 1980s, emphasises the *network* structure of the relationship between different innovation processes, organisations and sectors. For NIS scholars, innovation policy must focus on the system as a whole, and enable synergies between the different components of the system. Furthermore, because the success of a given institution cannot be understood without understanding its function within the broader political-economic innovation system, the 'same' policies or practices may have different impacts depending on the broader national ecosystem into which they intervene.

Freeman (1987, 1995) illustrates these points with a series of case studies from Japan, Russia, Brazil, South Korea and other economies. Moreover, as the NIS approach has gained prominence, scholars have extended it beyond the context of the nation state. Cooke et al. (1998) criticise the restriction of the first wave of NIS scholars to the national level of analysis, and advocate a focus on *regional* systems of innovation. Etkowitz and Lydesdorff (2000) return to the traditional organisational trio of government, university and firm to emphasise the complex interconnections between the three, with the metaphor of the 'triple helix'.

At the same time, a group of scholars informed by Gibbons et al.'s 1994 *The New Production of Knowledge* have argued that the major failure of the linear model was its inability to accommodate a structural transformation of knowledge production. For Gibbons et al. the linear model was once a more adequate account of the relationship between science and innovation, but the mode of knowledge production that the linear model captured - which Gibbons et al. refer to as 'Mode 1' - has since been supplanted by an alternative 'Mode 2' (cf. Nowotny, Scott, and Gibbons 2003). Where Mode 1 knowledge production emphasised pure science, Mode 2 emphasises 'contexts of application'. Where Mode 1 made heavy use of hierarchical organisational structures, Mode 2 emphasises 'heterarchical' forms of organisation. And where Mode 1 focused on narrow academic disciplinary boundaries, Mode 2 is transdisciplinary, valuing communication both between academic disciplines and between the academy and other institutions. Like NIS scholars, therefore, Gibbons et al. emphasise the importance of understanding the network structure of the entire innovation economy.

Finally, the 'open innovation' literature (Chesbrough, Vanhaverbeke, and

West 2006; Gassmann, Enkel, and Chesbrough 2010; Chesbrough, Vanhaverbeke, and West 2014; Bogers, Chesbrough, and Moedas 2018) has explored the importance of information sharing between firms and other entities, through public-private partnerships, breaking down research and development silos, collaborative product design, creation of platforms for knowledge pooling, and the incentivisation of 'open' sharing of research - including scientific findings and data. For Bogers et al. (2018: 6), this open innovation paradigm aligns closely with 'open science' approaches, as in the European Commission's current three overarching innovation policy goals: "Open Innovation, Open Science and Open to the World" (European Commission 2016: 7). Like scholars working within the 'National Innovation Systems' and 'New Production of Knowledge' paradigms, open innovation researchers tend to emphasise the importance of the innovation ecosystem, and the potentially beneficial reciprocal, rather than simply linear, relationships between its different components.

All of these approaches at times position themselves against a traditional linear model of the relationship between science and economic output - but none of them rejects the idea that basic scientific research is often a key input of the production process. In this sense, these alternative approaches are all compatible with a 'weak' version of the linear model (cf. Balconi, Brusoni, and Orsenigo 2010: 7). These alternative approaches all emphasise, however, that the causal chain from basic research to product innovation has to be located within a more complex understanding of the innovation process. The question of how best to understand - and influence - that broader innovation ecology is the subject of ongoing science policy debates.

In this thesis I am concerned with one particular subset of those debates - those focused on the value and function of 'open science'. In the remainder of this chapter, I will therefore outline the concept of 'open science'. In the next section I argue that, although scholars and advocates often talk about 'open science' in general, the term captures a range of quite different classes of institutional proposal. I distinguish four different senses of 'open science' and discuss which of these will be the primary focus of the remainder of the thesis.

1.2 Four meanings of 'open science'

There is no single, consensus definition of 'open science'. Michael Nielsen defines the concept as "the idea that scientific knowledge of all kinds should be openly shared as early as is practical in the discovery process" (Gezelter 2011); Neylon and Wu as the attempt "to open up the details of the scientific enterprise so that a wider range of people can participate" (Neylon and Wu 2009: 2). For Bartling and Friesike, open science aims to achieve "a scientific culture that is characterized by its openness" - a culture in which "[s]cientists share results almost immediately and with a very wide audience" (Bartling and Friesike 2014b: 10).

Further, there is no consensus over how consequential the institutional changes associated with open science are likely to be. For some scholars and advocates, open science potentially represents a "second scientific revolution" (Bartling and Friesike 2014b); for others, it represents little more than a defence of traditional scientific practice (Nelson 2004). To some extent this lack of consensus is attributable to substantive disagreements over how to model and evaluate open science. But to some extent it is attributable simply to the fact that different scholars are using the term 'open science' to refer to different things.

In this section I will attempt to cut through some of this potential confusion, by distinguishing four distinct meanings of the phrase 'open science', and by analysing four different classes of institutional proposal that are captured under this heading.[2] These four meanings - and the terms with which I will refer to them - are:

1. Free public use of the findings of scientific research - 'open findings';
2. Free public access to published scientific papers, reports, etc., without

[2]The typologisation of open science proposed in this section is novel, and other scholars propose alternative categories which also provide insight into the varieties of 'open science'. For Watson (2015) there are "six commonly accepted pillars of open science: open data, open access, open methodology, open source, open peer review and open education" (Watson 2015: 1). Masuzzo and Martens propose four 'pillars' of open science: open data, open code, open papers, and open review (Masuzzo and Martens 2017: 3). Fecher and Friesike schematise open science initiatives in terms of five schools of thought: democratic, pragmatic, infrastructure, public, and measurement (Fecher and Friesike 2014). The typologisation I propose here is not incompatible with any of these alternative frameworks - however, I suggest that my schema provides a useful window onto the issues that are central to this thesis.

payment or institutional affiliation - 'open access';

3. Broad public sharing by researchers of the intermediate inputs of scientific research, such as data, computer code, research instruments, etc. - 'open inputs';
4. Expansion of the range of social actors who participate in scientific research itself - 'citizen science'.

Of the phrases I adopt to denote these categories, *open access* and *citizen science* are standard terms. What I refer to as *open findings* is typically referred to, in the literature, as simply 'open science' - the potential for terminological confusion here is large, hence my adoption of an alternative phrase. What I refer to as *open inputs* is typically denoted, in the literature, in terms of the specific class of input the author has in mind - e.g. *open data*; *open source*.[3] I find the broader term *open inputs* useful in capturing this entire class of institutional proposals, though I will also often refer to open data, etc. as appropriate.

In the next four subsections I unpack each of these categories of open science in turn. In section 1.3, I then draw on these discussions to foreshadow themes that will recur through the remainder of the thesis.

1.2.1 Open findings

Open findings, as I will use the term in this thesis, denotes the idea that the *content* of scientific findings - scientific findings themselves, rather than their expression in specific publications - are available for free public use and reuse. 'Open findings' thus distinguishes scientific content that is freely and publicly available from content the use of which is restricted by means of either secrecy - e.g. state or trade secrets - or enforceable intellectual property law. In economists' terms, 'open findings' captures the idea that science is, or should be, a public good.

'Open findings' is an important part of the sense in which the term 'open science' is used by many historians of science - for example, by Paul David in his work on the norms of modern scientific inquiry (David 1998a). For David, as for other historians of science (e.g. Eamon 1985; Shapin and Schaffer 1985), one of the characteristic features of modern science is the willingness of the

[3]Though see Wagner (2010) for an example of the usage I adopt here.

producers of scientific knowledge to make that knowledge available for anyone to use as they see fit.

For many theorists of science, this public character of scientific knowledge arises naturally from the economic properties of knowledge itself. In his famous discussion of the governance structures best suited to the regulation of invention, for example, Thomas Jefferson wrote that an idea has curious properties that intrinsically orient it towards general public dissemination:

> the moment it is divulged, it forces itself into the possession of every one, and the receiver cannot dispossess himself of it. Its peculiar character, too, is that no one possesses the less, because every other possesses the whole of it. He who receives an idea from me, receives instruction himself without lessening mine; as he who lights his taper at mine, receives light without darkening me. That ideas should freely spread from one to another over the globe, for the moral and mutual instruction of man, and improvement of his condition, seems to have been peculiarly and benevolently designed by nature, when she made them, like fire, expansible over all space, without lessening their density in any point, and like the air in which we breathe, move, and have our physical being, incapable of confinement or exclusive appropriation. (Jefferson 1854: 180-1)

For David, however, the idea that ideas should spread freely by nature's design is not, despite appearances, intrinsic to the innate properties of conceptual or intellectual resources - rather, it is a social accomplishment. In David's words:

> Although the conceptualization of science as the pursuit of "public knowledge" now seems to many a natural, even a primitive notion, it is in reality a complex social construct. (David 1998a: 15)

In common with many other historians of science, David argues that this construct can be dated from the 'epistemological revolution' associated with late 16th and 17th century experimentalist and rationalist intellectual projects. This revolution, in addition to transforming the methods with which the pursuit of knowledge was carried out, also involved:

> a break from the previously dominant regime of "secrecy in the pursuit of nature's secrets". (15)

In Eamon's words:

> Medieval science, conditioned by a rigid social hierarchy and by a political and religious outlook that encouraged withholding secrets of nature from the "vulgar", also failed to develop an outlook that promoted openness. Only in the sixteenth and seventeenth centuries, with the emergence of new technology, new institutions for the promotion of scientific activity, and institutional mechanisms to protect the interests of discoverers, did the conception of science as "public knowledge" take form. (Eamon 1985: 321-2)

For these historians of science, public findings should therefore be seen as an achieved institution, and the achievement of that institution is an appropriate object of social-scientific study.

1.2.2 Open access

Open access, as the term is generally used - and as I use it in this thesis - refers to the broad, free public availability of scientific *publications*: the lowering of publication access barriers, to make scientific texts easy to read and to use for a very large, potentially global, audience. In the words of the 'Budapest Open Access Initiative':

> The literature that should be freely accessible online is that which scholars give to the world without expectation of payment... By "open access" to this literature, we mean its free availability on the public internet, permitting any users to read, download, copy, distribute, print, search, or link to the full texts of these articles, crawl them for indexing, pass them as data to software, or use them for any other lawful purpose, without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself. (Chan et al. 2002)

More specifically, open access is often schematised along two dimensions: the free online availability of scientific publications; and the lowering of copyright claims over scientific publications, to permit their easier republication and re-use. Among open access advocates, these two senses of the term are often referred to as 'gratis' and 'libre' open access - in Peter Suber's adaptation of Richard Stallman's remark: "gratis OA is free as in beer, while libre

OA is also free as in speech" (Suber 2015; cf. Stallman 2015). In terms of their institutional consequences, libre open access involves transformation of copyright regulations; gratis open access involves transformation of publishing models.

For libre open access, the most prominent and widely adopted of the proposed institutional changes are 'open copyright' licenses, such as those made available by the non-profit organisation Creative Commons. In the words of Lawrence Lessig, one of the organisation's founders, creative commons licences allow authors to:

> select the freedoms you want to grant. You can say, "I want people to give me attribution or not," or you can say, "I want people to use this for commercial use or not," or you can say, "I want to allow people to modify this or not," or you can use what we call a "share alike" license that says, whatever freedoms you got from me, you have to pass on to someone else. (Lessig 2004: 11)

Creative commons copyright licenses therefore offer a range of different open access possibilities - as do the most notable alternative OA copyright proposals, such as OpenContent.org's 'Open Publication License' (OpenContent 1999).

Gratis open access, by contrast, does not require a significant change in copyright practices - but does require that publications be accessible without charge on the public internet. Gratis open access is the most commonly used sense of the term - in part because of its increasingly widespread adoption in the academic publishing industry. In a 2010 study of the growth of open access publishing, Laakso et al. tracked the growth of both open access journals and open access publications. At the starting date of their study, 1993, the authors found 247 open access articles, published in 20 open access journals; in the last year of their study, 2009, they found 191,851 articles, published in 4767 journals. This figure has continued to grow: in a 2018 study Piwowar et al. used oaDOI.org (now 'rebranded' as unpaywall.org) to estimate the proportion of published articles which are open access, finding that approximately 27.9% of DOI-assigned journal articles are open access, a proportion that has grown steadily since the mid-1990s (Piwowar et al. 2018: 10-12). As of October 2019 the Directory of Open Access Journals listed 13,785 journals (DOAJ 2019).

The most prominent of the open access publishing models are so-called 'gold' and 'green' open access.[4] 'Gold' open access is publication of research papers in journals that are freely available on the public web. Two funding models for Gold open access have been widely adopted: on the one hand, subsidy by, most commonly, an affiliated institution; on the other hand, payment by authors, or their institutions, to cover the costs associated with publication. 'Green' open access, by contrast, involves publication of research papers in traditional 'closed' journals, coupled with the archiving of the paper in a publicly accessible research repository (Björk et al. 2014). Almost all academic journals now permit such public archiving, with variation in whether post-publication or pre-publication versions of an article may be archived; whether and what embargo periods are required; and what repositories are and are not permitted. Finally 'dual' open access refers to journals that host both open and closed articles, and where authors can pursue either green or gold access options.

These are the major forms of open access publication model. However, it is also worth noting an alternative set of related proposals: the overlay journal (Ginsparg 1997), decoupled journal (Priem and Hemminger 2012), or deconstructed journal (Smith 1999).

In 'overlay journals' scholarly content is hosted and published in gratis open access institutional repositories, but curated by a 'journal' that functions simply as an aggregator and certifier of quality (Ginsparg 1997; Rodriguez, Bollen, and Sompel 2006; Moyle and Lewis 2008). In the 'deconstructed' or 'decoupled' journal, the different functions of a journal are separated still further, aspiring to disaggregate *all* the different elements of the academic publication process (Smith 1999; Priem and Hemminger 2012). In Smith's words, the deconstructed journal builds on the insight that most of the activities involved in the production of a journal "are independent... and therefore there is no obvious need for these roles to be carried out, or coordinated, by a central organisation (i.e. the publisher)" (Smith 1999: 82). Overlay journals have been trialled (Rodriguez, Bollen, and Sompel 2006; Moyle and Lewis 2008), and in 2016 the mathematician Timothy Gowers launched the journal *Discrete Analysis* using the overlay model (Ball 2015), but the uptake of the overlay journal approach has nevertheless been limited

[4]The discussion below will focus on the academic journal - but many of the same principles apply to other forms of academic publication. See Willinsky (2005) for a comprehensive survey of different elements of the open access movement.

- these remain underexplored institutional possibilities.

1.2.3 Open inputs

Open inputs denotes a large class of institutional proposals that aim to make easily publicly accessible the *intermediate inputs* of scientific production. To the extent that research findings and scientific publications are also inputs into scientific production, 'open inputs' can be seen as a category that encompasses the previous two. But I will be using it to refer to those inputs that only or primarily exist *as* inputs - the open sharing of scientific materials that would otherwise not be available outside the research process.

Here I will discuss some of the most important of these inputs: open source; open data; open archiving; open research protocols; and open notebook science.

Open source, as the term is most commonly used, is a broad movement producing software in which the source code is non-proprietary - that is to say, publicly available for modification and reuse. The open source movement has been the subject of considerable study by economists and social theorists (e.g. Lerner and Tirole 2000; Benkler 2002, 2006), in part because it appears to stand in tension with much traditional economic theory: open source production often operates without either the forms of hierarchical organisation, or the private property incentive structures, associated with much economic activity. For this reason, open source production has become an iconic representation of "commons" or "peer production". In the context of open science, however, 'open source' refers more narrowly to the practice of making publicly available online the source code used in scientific production (González 2005; Easterbrook 2014).[5]

Open data refers to the practice of making the data that is the basis of scientific conclusions available online for general use by other researchers (Molloy 2011). Open data has been embraced by many governments and NGOs (Huijboom and Broek 2011; Attard et al. 2015; Nugroho et al. 2015; European Commission 2016). Likewise, some journals now require the sharing of data as a condition of publication (Wicherts et al. 2006; Naughton and Kernohan 2016; Hardwicke et al. 2018). In addition, there is a growing movement oriented towards encouraging data sharing as a scientific norm,

[5]See Wilinsky (2005) for a discussion of connections between the open source, open access and open science projects.

independent of institutional requirements (Vision 2010: 330; Munafo et al. 2017; Corti et al. 2019).

'Open data' does not just denote the online sharing of research-created data; it also captures the digitisation of existing data sources - a practice that is increasingly common. As Bolick puts it, with reference to the discipline of history: "Conducting historical research with primary sources has been traditionally reserved for scholars who had the means to physically visit historical sites, libraries, and archives" (Bolick 2006: 122). Increasingly, however, such sources are being digitised and made publicly available, a development that has "democratized historical research" (122) by creating another form of open data repository.

Another important form of open data is *open research protocols*. Debates over the 'replicability crisis' in the sciences - which I discuss in more detail in chapter seven - have increasingly focused researchers' attention on the risks associated with the 'file drawer problem' (Rosenthal 1979) and 'researcher degrees of freedom' (Simmons, Nelson, and Simonsohn 2011). For those worried about these issues, published research may not accurately represent the object of its study because of systematic institutional incentives to withhold inconvenient findings from publication, and to adopt methodological approaches that tend towards the production of false positive results (Earp and Trafimow 2015). One suggestion to improve these incentive structures is the norm of open publication of research protocols before the research itself begins, to allow scholars to assess the final published literature against initial research plans (Goldacre and Gray 2016; Lucey, Clark, and Glasziou 2017). To this end, there are increasing efforts among practitioners to establish both the technical infrastructure and the norms of practice to facilitate widespread preregistration of research protocols (Nosek et al. 2018).

Finally, some scientists advocate an even more comprehensive sharing of the intermediate inputs of scientific production, by sharing not just protocols, data or research tools like code, but the notes that are collected as part of day-to-day scientific practice. This shift has been made technologically feasible by the increasingly widespread use of 'Electronic Lab Notebooks' (ELNs) (Phillips 2006). The use of ELNs makes possible 'open notebook science', which aims to extend such data sharing beyond the immediate research team, by "making the entirety of one's laboratory notebook and all associated raw data public in as close to real time as possible" (Stafford 2010: S21). Although Bradley and some other researchers (Williams 2008: 503) have experimented with open notebook science, however, it remains a

niche practice.

I will return to the discussion of open inputs in more detail in Part III.

1.2.4 Citizen science

Finally, *citizen science* denotes science carried out by amateurs. The term was coined by Alan Irwin in his 1995 book *Citizen Science*. In Irwin's words: "'Citizen Science' evokes a science which assists the needs and concerns of citizens... At the same time, 'Citizen Science' implies a form of science developed and enacted by citizens themselves" (Irwin 1995: xi). In common with most scholars in the field, I adopt the second of Irwin's definitions here.

Amateur science is not a new phenomenon: for much of the history of science, scientific research has primarily been carried out by amateurs, and even in recent decades many major forms of science would not have been possible without the contributions of non-scientists. As Bonney et al. put it:

> Much of our current understanding about our natural environment... is derived from data that have been collected, transcribed, or processed by members of the public. (Bonney et al. 2014: 1436)

But as with the other forms of open science discussed above, recent technological changes have led to a fundamental change in the possibilities associated with citizen science. The development of internet technology, together with broader access to potential data collection tools like smartphones, has enabled a very substantial increase in citizen science participation in all stages of the research process.

Thus, for example, amateur scientists are increasingly involved in data collection. The UK National Biodiversity Network, for example, collates and standardises wildlife data collected by citizens, and makes the data available online for researchers to use (NBN 2019). eBird.com enables US birdspotters to create a very large database of bird observations (eBird 2019). The US Community Collaborative Rain, Hail and Snow Network collects and aggregates meteorological data from volunteers (CoCoRaHS 2019). In Silvertown's words: "Almost any project that seeks to collect large volumes of field data over a wide geographical area can only succeed with the help of citizen scientists" (Silvertown 2009).

Citizen science also has an increasing role in analysing quantities of data the scale of which exceed the capacity of small research teams. One striking

example of this is Galaxy Zoo, a project which takes images of galaxies automatically collected as part of the Sloan Digital Sky survey, and asks members of the public to classify them using a simple online web interface (Lintott et al. 2008; Raddick et al. 2010). Galaxy Zoo has been enormously successful in mobilising the manpower of citizen volunteers to code a vast quantity of data - a process that has not only generated important statistical findings, but discovered novel astronomical objects (Lintott et al. 2009; Nielsen 2012).

It is intuitive that data collection and simple data coding can, in principle, be carried out by non-professionals. Less intuitive is how scientific 'problem solving', of the kind that apparently requires extensive background knowledge of the scientific principles governing the objects under analysis, is amenable to amateur contributions. For this reason one of the most notable innovations in citizen science is the use of 'gamification' to 'embed' specialist scientific knowledge within a software environment, allowing non-specialists to problem-solve in ways that would normally require scientific training.

One example of this practice is Phylo - a game which aims to crowd-source solutions to the problem of aligning multiple sequences of DNA (Kawrykow et al. 2012). The innovation of Phylo "is to reduce the human computing part to a casual game, a puzzle, in order to broaden the spectrum of participants and collect the computing power generated by regular, non-scientist gamers" (2). A related use of gamification to crowdsource solutions to hard scientific problems is FoldIt, which invites players to manipulate protein structures to solve real-world modelling problems (Khatib et al. 2011). Both FoldIt and Phylo have enabled casual, non-scientist players to make significant contributions to ongoing scientific research projects.

Alongside these involvements in the research process itself, citizen science has also contributed to the funding of science, with the development of science-specific crowdfunding platforms like experiment.com and scifundchallenge.org (Wheat et al. 2013; Siva 2014). Finally, the term is increasingly used to encompass research designed and carried out in full by amateurs (Wohlsen 2011).

1.3 The focus of the thesis

In the previous section, I outlined four different meanings of 'open science'. In later chapters of the thesis, I will discuss in more detail debates over the strengths and weaknesses of these institutional proposals. In this final section

of chapter one, I will foreshadow those discussions, by surveying the three main issues at stake in the debates over open science. This will then allow me to state more clearly the focus of the thesis.

All of the institutional proposals discussed in the previous section are matters of some controversy, with advocates and critics of each. Speaking schematically, three overarching issues are at stake in these debates.

First: the question of the role of science in the broader economy. Some of the issues at stake in this debate I have already discussed in section one above. In general, the dimension of open science most relevant to this issue is open findings, because the way in which scientific findings feed into other dimensions of economic production is most directly influenced by the availability for reuse of those findings.

Second: debates around the *efficiency* of science. Here debates most often focus on the problem of open inputs. Scientists who share the intermediate inputs of scientific production reduce duplication of effort and thereby increase the efficiency of scientific production; those who refuse to share intermediate inputs correspondingly reduce scientific efficiency. I will return to this problem in more detail in chapter two.

Third: debates around the accuracy, or *credibility*, of science. Critics of some open science institutional proposals charge those proposals with undermining the reliability of scientific findings: for example, critics of elements of the open access movement argue that many open access publications have lower or absent peer review standards, compared to their closed access rivals; critics of citizen science argue that amateur scientists are often less skilled, and their findings therefore less robust, than professional scientists; etc.

This very schematic summary of some key debates over open science is intended merely to foreshadow the more extensive treatment of these issues later in the thesis. However, it also allows me to make clearer, in a preliminary way, the focus of later chapters.

Of the four classes of open science institutional proposal, the thesis will focus primarily on *open inputs*; of the three areas of institutional debate, the thesis will focus on the problems of *efficiency* and *credibility*. More precisely, the topic of the thesis can be stated as follows: what role, if any, can open science institutions play in increasing the *efficiency* of scientific production, via the increased sharing of *intermediate inputs*, subject to the constraint of retaining or increasing the *credibility* of scientific output?

In order to unpack this further, and understand how best to approach these problems, I need to draw on a range of resources in the economics of

science. In the next chapter I will therefore give an overview of the economics of science, with particular emphasis on 'the new economics of science', 'the economics of scientific knowledge', and the formal toolkit of agent-based modelling of scientific research dynamics.

Chapter 2

The economics of science

2.1 Introduction

In chapter one I gave an overview of the different senses in which the term 'open science' is used. In later chapters I will compare some of these open science proposals with more traditional scientific practice. In this chapter I lay the groundwork for that discussion, by reviewing the theoretical literature in the economics of science. Just as, in the last chapter, I discussed how the mid-twentieth century linear model was challenged by a range of more recent analyses of the role of science in the innovation economy, here I will discuss how mid-twentieth century neoclassical economics of science has been challenged along several distinct dimensions.

In section 2.2, I discuss the economics of science associated with the postwar neoclassical theorists, the formal starting-point for much economic theory in the second half of the twentieth century. In the following four sections I then discuss four different departures from and developments of that neoclassical starting point. In section 2.3, I discuss the shift at the *macroeconomic* level of analysis from exogenous to endogenous growth theory - and the central role given to knowledge production within the latter. In section 2.4, I discuss the development at the *microeconomic* level of a more game-theoretic and institutionalist approach to scientific production, associated with the 'new economics of science'. In section 2.5, I discuss the application of economic tools to epistemological questions more traditionally associated with the philosophy of science, under the heading the 'economics of scientific knowledge'. Finally, in section 2.6, I discuss the recent increasing applica-

tion of the tools of agent-based modelling to the analysis of the dynamics of scientific research. These last three approaches, in particular, will provide resources that I will return to and build upon in later chapters.

2.2 Early neoclassical economics of science

The starting point for much modern discussion of the economics of science is the postwar neoclassical approach to knowledge as a public good. For the economic theorists of this neoclassical tradition, the formal apparatus of equilibrium analysis and social choice theory could be used to analyse both the *positive* question of how economic resources *would* be allocated under different circumstances, and the *normative* question of how economic resources *should* be allocated to maximise total social welfare.

One of the fundamental distinctions informing this analysis was the distinction between *public* and *private* goods. In an influential paper on 'The pure theory of public expenditure' (1954) Paul Samuelson distinguished between

> ordinary *private consumption goods*... which can be parcelled out among different individuals... and *collective consumption goods*... which all enjoy in common in the sense that each individual's consumption of such a good leads to no subtraction from any other individual's consumption of that good... (Samuelson 1954: 387)

Later theorists were to disaggregate this distinction into two more fundamental distinctions: between *rivalrous* and *non-rivalrous goods* and between *excludable* and *non-excludable* goods (see e.g. Buchanan 1965; Ostrom 1990). In this terminology, the consumption of *rivalrous* goods reduces the quantity available for others to consume; the consumption of *non-rivalrous* goods does not. It is easy to restrict access to *excludable* goods, but hard to restrict access to *non-excludable goods*. These categories therefore give a familiar two-by-two matrix of types of good:

	Rivalrous	Non-rivalrous
Excludable	Private goods	Club goods
Non-excludable	Common-pool resources	Public goods

In terms of this schema, Samuelson was concerned only with *private* and *public* goods - his framework does not consider the role of club goods or common pool resources. I will return later in the thesis to the question of how the analysis of scientific production changes once we take fuller account of these categories.

For Samuelson, then, production was restricted to the production of private (excludable, rivalrous) consumption goods, and collective (non-excludable, non-rivalrous) consumption goods. Moreover for Samuelson this dichotomy between types of goods corresponded to a dichotomy between the two principal institutional structures governing economic production. Private consumption goods were best produced by firms trading in the *market*, and collective consumption goods were best produced by the *state*.

The argument for this dichotomous approach to institutional structures was grounded in the formal apparatus of welfare economics. In an economy dealing purely with private goods the private self-interested utility- and profit-maximising actions of individual consumers and firms will, via "market catallactics" (389) produce the aggregate social optimum:

> on the one hand perfect competition among productive enterprises would ensure that goods are produced at minimum costs and are sold at proper marginal costs, with all factors receiving their proper marginal productivities; and on the other hand, each individual, in seeking as a competitive buyer to get to the highest level of indifference subject to given prices and tax, would be led as if by an Invisible Hand to the grand solution of the social maximum position. (388)

This situation breaks down, however, when collective or public goods are included in the model. Public goods disrupt the smooth welfare-maximising functioning of the perfectly competitive market economy for two reasons.

First, because public goods are intrinsically non-excludable, their production and consumption will always potentially involve *externalities* the (positive or negative) value of which will not be captured in private market exchange. Thus while there is a socially optimal level of public good production, there is no reason to expect private production to correspond to this social optimum. For Samuelson, therefore, public expenditure must step in, to make up the difference between the private production of a public good and its socially optimal level of production.

Second, however, there is a fundamental problem of determining the socially optimal level of public good production. The perfectly competitive market in the neoclassical model is not just a mechanism of coordination but also of information transmission (cf. Hayek 1945). While a socially optimal level of public good production must exist, determining that level requires information about consumers' preferences. Obtaining this information is not just a practical problem about data collection or the aggregation of heterogeneous subjective preferences; it is also a problem of incentivising the truthful revelation of those preferences. In a perfectly competitive market economy of private goods, consumers have no incentive not to represent their preferences accurately in exchange. In communicating with a social planner about their preferences around public goods, by contrast, there are strong incentives to *mis*represent those preferences. In Samuelson's words:

> it is in the selfish interest of each person to give *false* signals, to pretend to have less interest in a given collective consumption activity than he actually has, etc. (Samuelson 1954: 388-9)

These two issues, of *externalities* and *information coordination*, are central to the economics of public goods that developed from this starting-point. But they are not the only obstacles to the efficient production of knowledge as a public good. In a 1962 paper, Kenneth Arrow raises an additional issue around the production of scientific knowledge (Arrow 1962): the problem of resource allocation associated with risky investment. Arrow notes that scientific and technological innovation is an intrinsically risky investment, since not just the market value but the very content of such output cannot be known ahead of time. Furthermore, for Arrow this issue is an expression of a broader problem, which he calls:

> a fundamental paradox in the determination of demand for information; its value for the purchaser is not known until he has the information, but then he has in effect acquired it without cost. (615)

For Arrow this is a profound difficulty with the attempt to create a market for private exchange of information, and provides further warrant for the conclusion that "[t]here is a strong case for centralized decision-making under these circumstances" (616) - but this "paradox" of information production

and exchange has ramifications for *any* economic model that aims to analyse the role of information in the economy - not just models that focus on market demand.

Despite these issues, the mid-twentieth century neoclassical theorists were united in the idea that the public production of knowledge as a public good was the best 'baseline' model for thinking about science. This orientation is perhaps best expressed by Richard R. Nelson's (1959) paper 'The simple economics of basic research'.

In that paper, Nelson draws a distinction between basic and applied research. Applied research is oriented toward the production of a specific invention or good - Nelson gives the examples of "Edison's attempt to develop an incandescent lamp and Goodyear's attempt to improve the characteristics of rubber" (300). Basic research, by contrast, may result in applications, but does not have those applications as its motive. For this reason, Nelson argues, fundamental discoveries that are transformative of technological possibilities more commonly arise from basic than from applied research. Nelson gives the examples of the work of Maxwell and Hertz that enabled the development of radio communication:

> It seems most unlikely that a group of scientists in the mid-nineteenth century, attempting to develop a better method of long-range communication, would have developed Maxwell's equations and radio or anything nearly so good. (301)

Building on these remarks, Nelson adds an additional item to the collection of problems with privately funded research. There are two obvious ways in which private producers of research can receive income from their research findings. On the one hand, they can use their findings to produce goods for sale on the market; on the other hand, they can sell the findings on the market directly. The former is only possible for research with direct practical application, and therefore typically cannot be used to incentivise basic research. The latter is only possible if intellectual property law allows the ownership of research findings to be asserted and enforced. However, asserting intellectual property over research findings is highly inefficient in terms of broader social outcomes. One of the reasons knowledge is so important to economic growth is that the 'same' piece of knowledge can be an input for a very wide range of different production processes; asserting intellectual property over scientific findings prevents this broad social use of scientific knowledge. In Nelson's words:

> there is a basic contradiction between the conditions necessary for efficient basic research - few or no constraints on the direction of research with full and free dissemination of research results - and full appropriation of the gains from sponsoring basic research in a competitive economy. (305)

For Nelson it is therefore logical that the majority of basic research would be publicly funded. More specifically, basic research is only likely to be adequately funded by private companies in an economy where those companies have strong monopoly control over very large sectors of the economy, such that even investment in basic research can be recouped via its impact on applied research (cf. Nelson 1962). Absent such monopoly control, Nelson argues, the public sector must provide the incentive for much basic research to take place:

> If we want to maintain our enterprise economy, basic research must be a matter of conscious social policy. (306)

In chapter one, I described the linear model of the connection between science and innovation, and the association between this linear model and 'top down' approaches to planned scientific production. Here, in Nelson's early work, we can see the paradigmatic expression of that understanding of scientific research in the formal apparatus of neoclassical economics.

In summary, then, the post-war neoclassical theorists identify science as a *public good*. Public goods, they argue, always generate positive externalities, and are therefore intrinsically under-produced by market mechanisms. For this reason, their production should be funded or subsidised by taxpayer funded state activity. Identifying the socially optimal level of state subsidisation is a difficult problem for two reasons. First, it is impossible to reliably estimate preferences for public good production, because there are intrinsic disincentives for private economic actors to truthfully disclose those preferences. And second, it is impossible to reliably estimate the value of information prior to the possession of that information. Nevertheless, despite these obstacles to estimating the appropriate level of public subsidy for scientific production, it is clear to the post-war theorists that the level of public subsidy should be substantial.

This was a simple, clear framework, which informed much of the economics of science for many decades, and continues to exert considerable influence in more recent work (e.g. Stiglitz 1999). Nevertheless, in the later

decades of the twentieth century, economic theorists broke with this approach in several ways.

In the remainder of this chapter, I will look at four important traditions that both build upon and depart from this early foundational work in the economics of science. In section 2.3, I look at the shift from *exogenous* to *endogenous* growth theory, and the repercussions this shift had for how the knowledge production process is conceptualised by economic theorists. In section 2.4, I look at the institutional turn in the 'new economics of science', and the impact this has had on our understanding of the micro-incentives of scientific production. In section 2.5, I look at the 'economics of scientific knowledge', and the way in which this relatively new set of approaches has drawn economists' attention to some of the peculiar features specific to scientific knowledge as a good. Then in section 2.6, I look at the toolkit of agent-based modelling, and how it has recently been used to study the dynamics of science as an internally heterogeneous complex system.

2.3 Knowledge and growth theory

It is common to schematise growth theory into two waves: 'old', exogenous, and 'new', endogenous, growth theory. In this section I will quickly survey both, with an emphasis on the latter, and in particular on the role of knowledge in economic growth.

The point of departure for much modern growth theory was the Keynesian macroeconomics that emerged and became dominant in the discipline from the late 1930s. Although Keynesian macroeconomics was primarily focussed on short term changes in output and employment, the tools of Keynesian analysis were also adopted to the analysis of long term growth (Harrod 1939; Domar 1946). In the 'Harrod-Domar model' the growth rate was determined by a set of exogenously given parameters; furthermore, the economy as modelled had no intrinsic orientation toward stable growth at high employment rates. For Harrod and Domar economies would often be expected to exhibit severe unemployment or severe under-utilisation of capital, with no intrinsic self-correcting or self-equilibrating mechanism.

For a generation of economists preoccupied with the Great Depression and its aftermath, this orientation to growth theory seemed compelling. In the postwar years, however, the 'pessimism' of the Harrod-Domar model seemed increasingly at odds with the stability of the major industrial economies;

furthermore, the theoretical foundations of the approach were challenged by the increasing mathematical sophistication of the emerging neoclassical paradigm. In the mid-50s, therefore, a new neoclassical growth model was developed (Solow 1956; Swan 1956), which soon came to be regarded as the baseline model for modern growth theory.

In the Solow model, the Harrod-Domar model's tendency to disequilibrium is removed by endogenising one of the key variables - the output-capital ratio (cf. Solow 1994). In place of the Harrod-Domar crisis-prone economy, Solow proposes a model of the economy that will, *ceteris paribus*, move towards a steady state in which output per capita is constant. For Solow, shifts in the investment or depreciation rates will shift the steady state, but will not influence long term growth; thus in the Solow model the growth rate itself is a function of, on the one hand, population growth and, on the other hand, increases in total factor productivity. Since it is clear that population growth can in itself explain only a small proportion of historical and national growth rate differentials, economists' attention soon came to focus on the analysis of changes in total factor productivity as the key to understanding long term economic growth.

In his influential 1957 paper 'Technical change and the aggregate production function', Solow underlined the importance of this focus, by seeking to statistically distinguish "shifts of the aggregate production function from movements along it" (Solow 1957: 320). The 'Solow residual' - the proportion of growth that cannot be accounted for through capital accumulation or labour growth - then became a central focus for economic analysis. Economists asked two questions. First: in what, exactly, did the residual consist - technological change, institutional or organisational change, and/or human capital growth - and in what proportions? Second: how could changes in this 'residual' - changes, that is to say, in total factor productivity - be incorporated endogenously into models of economic growth, rather than assumed as an exogenously given change in an exogenously specified parameter?

Endogenous growth theory aimed to answer these questions. The unifying feature of this approach, in addition to its attempt to model the growth of Solow's residual endogenously, is the idea that the residual should be understood as knowledge in some sense. The differences among the different strands of endogenous growth theory can therefore usefully be understood as differences in how to understand, and model, knowledge and its production.

The first wave of endogenous growth theory aimed to endogenise growth by seeing growth in knowledge as an 'accidental' output - a positive exter-

nality - generated by the elements of production already analysed within the neoclassical model. In his 1962 paper on 'learning by doing', for example, Arrow assumed that investment in capital production generated experience, as workers develop new skills by interacting with the fresh stock of physical capital; this experience in turn increases the productivity of capital, leading to a 'virtuous circle' of capital investment and increases in total factor productivity.

This 'positive externalities' approach was important as the first major effort to endogenise the growth of knowledge, but it was widely seen as inadequate - largely due to its neglect of the 'casual empiricist' observation that many of the increases of knowledge in modern economies are the result of deliberate actions on the part of economic actors taken with increased knowledge as their goal, rather than accidental byproducts of other economic activity. With this point in mind, the second wave of endogenous growth theory - while often incorporating 'learning by doing' as a subsidiary feature of its models - sought to analyse increases in knowledge as a product of intentional action by economic agents.

This more recent endogenous growth theory is a diverse field, but two major research programs can be distinguished. On the one hand human capital approaches, exemplified by Lucas (1988), see knowledge as embedded in human workers, and as produced by training. On the other hand technological approaches, exemplified by Romer (1990, cf. 1986), see knowledge as embedded in technological innovations, and produced by firms engaged in market-oriented research and development.

In Lucas' human capital model, workers in the economy may choose to divide their time between work, on the one hand, or human capital accumulation in the form of training, on the other. Work generates immediate income, while human capital accumulation requires forgoing present income in exchange for increased future income. Workers' decisions about what proportion of time to spend on human capital accumulation are influenced by their 'patience' - the discount rate - and by the gains in productivity associated with increased human capital. Long term economic growth is in turn explained by growth in human capital, and thereby partially endogenised.

For Romer, by contrast, knowledge is acquired via private firms' research and development process, and embedded in technological innovations. Romer's 1990 model proposes an economy with three sectors: a competitive research sector, which makes use of the existing stock of knowledge to create patentable product designs; a monopolistic intermediate goods sector, which

purchases designs to produce capital goods; and a competitive final output sector, which makes use of capital goods to produce consumption goods.

The core feature of Romer's model is the idea that new knowledge, when produced, is nonrival but *partially* excludable. A new design produced by the research sector can be patented and sold to firms who thereby gain a monopoly over the production of goods of that design. At the same time the knowledge embedded in the design is a public good, and can be used by the entire research sector as an input into the production of future designs. On Romer's model, therefore, the production of knowledge is incentivised - and endogenised - by its partial excludability, and by the flow of revenue that results from that excludability, yet knowledge production also involves very substantial positive spillover effects. Research, on Romer's model, therefore involves a constantly growing inventory of possible designs, *and* a constantly growing 'reservoir' of public knowledge, both of which have positive impacts on economic growth.

This second wave of endogenous growth theory revolutionised the field, yet no consensus has yet emerged around which approach is best equipped to analyse the role of knowledge in the 'knowledge economy'. In the remainder of this section I will therefore discuss two significant alternative approaches, which both build upon the second wave of endogenous growth theory, while challenging some of its assumptions: Schumpeterean growth theory; and work that focuses on the role of political institutions in growth.

Schumpeterean growth theory, exemplified by Aghion and Howitt (1990), has emerged as an important extension of, and challenge to, Romer's analysis of knowledge growth through research and development. The human capital and technological change approaches treat growth as driven by a linear accumulation of knowledge; Schumpetereans, by contrast, see innovation as a process of "creative destruction" (Schumpeter 1994: 82-3) that is both stochastic rather than deterministic, and capable of reducing, as well as increasing, overall value.

In Aghion and Howitt's model, just as in Romer's, technological growth takes place through innovation in intermediate inputs, production of which is carried out in its own sector. However, where in Romer new intermediate inputs simply add to the stock of existing intermediate inputs, resulting in linear growth in available technologies, for Aghion and Howitt new technology may also render old technology obsolete. This means both that the growth path is less predictable than in earlier endogenous growth theories, and that the incentives for innovation are more ambiguous: in some scenarios,

Aghion and Howitt argue, the negative impact of innovation on existing product monopolies may disincentivise innovation altogether. Schumpeterean approaches therefore require a more nuanced approach to the gains, and losses, from innovation than Romer's 'knowledge accumulation' model.

Finally, a range of recent theorists have also emphasised the importance of institutions in explaining growth rate differentials between societies. In recent work, North, Wallis and Weingast use a metaphor that resonates with the topic of this thesis to distinguish between 'open access' and 'limited access' societies (North, Wallis, and Weingast 2006, 2009). In this framework 'limited access' societies involve powerful social elites extracting rents from general economic activity by severely restricting access to the gains from both political and economic institutions. 'Open access' societies, by contrast, involve broader participation in both political and economic institutions.

A similar point, using different vocabulary, is made by Acemoglu and Robinson (2005, 2013), who distinguish between *extractive* and *inclusive* political and economic institutions. Acemoglu and Robinson aim to explain the emergence of specific institutions by analysing the balance of *de facto* power between different social groups. This analysis is, in turn, partly an 'endogenous' one, since differential political power can partly be explained by differential economic outcomes across society.

Such institutional approaches remain underdeveloped relative to the broader literature on endogenous growth theory. But they remind us that the decisions made by economic actors modeled by much growth theory take place within an institutional framework that cannot be taken for granted, and that the nature and transformation of that framework may itself be a major contributing factor in economic growth.

2.4 The new economics of science

The growth theory discussed in the last section aims to analyse the long term macroeconomic determinants of economic growth - but, for this reason, it does not typically analyse the specific institutions, the microeconomic incentive structures, or the local production processes involved in scientific research. From growth theory's macroeconomic perspective, scientific research is a 'black box' that for modelling purposes can be represented as a production function.

In the 1980s and early 1990s a range of researchers aimed to 'open the

black box' of scientific research, by examining the institutional structures that organised scientific research at a more micrological level. Most notably, Partha Dasgupta and Paul David (Dasgupta and David 1987; Dasgupta 1989; David 1994; Dasgupta and David 1994) developed a new approach to the incentive structure motivating scientific researchers. Dasgupta and David's work in turn drew on two principal sets of resources: the sociological theories of Robert K. Merton, and a game-theoretic approach to understanding institutional incentives. In the subsections below I examine each of these in turn.

2.4.1 Robert K. Merton's sociology of science

Within sociology Robert K. Merton is perhaps best known for his development and advocacy of a form of functionalism, exemplified in his 1957 work *Social theory and social structure*. But Merton also made major contributions to the sociology of science. Beginning in the early 1940s, Merton developed a new institutional approach to the study of science that served as the baseline approach for a generation of sociologists (Merton 1957a, 1961, 1968, 1973, 1988). Dasgupta and David's first contribution to the economics of science was the adaptation of this Mertonian approach to the discipline of economics.

While Merton's publications on science and the problems he addresses are varied, his principal contributions can be summarised in two points.

First Merton breaks down what he regards as the key norms informing scientists' world-view. For Merton:

> Four sets of institutional imperatives - universalism, communism, disinterestedness, organized scepticism - are taken to comprise the ethos of modern science. (Merton 1973: 270)

In this schema, universalism means that the "acceptance or rejection of claims entering the lists of science is not to depend on the personal or social attributes of the scientist" (270); rather, objective standards, which transcend nationality, class, religion, race, etc., determine the credibility of scientific claims.

Communism, for Merton - often restyled "communalism" or "communality" (e.g. Barber 1952: 130-132) - means that the findings of scientific research are "assigned to the community... property rights in science are whittled down to a bare minimum by the rationale of the scientific ethic." (273)

Merton therefore defines science as producing 'open findings' in the sense discussed in chapter one.

Disinterestedness, for Merton - although characterised as a norm - is more strictly a set of institutional checks and balances that can also often be internalised by working scientists: "the activities of scientists are subject to rigorous policing... The demand for disinterestedness has a firm basis in the public and testable character of science" (276). Merton's claim is that the mutual monitoring of scientists means that working scientists find it difficult to mislead the community for their own benefit, resulting in an internalised norm of disinterested research.

Finally, organised scepticism, for Merton, means that no claim is immune from critique. This applies to the claims of science, which must be internally tested; and also to claims made by figures and institutions outside of science, a fact which has "periodically involved science in conflict with other institutions" (277).

For Merton these four principles are the core normative framework which defines science as an institution, and within which scientific research is carried out.

Second, Merton analyses the *incentive structure* informing the decision-making of working scientists. For Merton, this incentive structure is not primarily economic. Professional scientists are motivated by income; but they are not motivated solely by income, and the mechanisms via which they can achieve higher incomes are themselves aligned with other more directly incentivised actions. Specifically, for Merton working scientists are also centrally motivated by *reputation*. The reputation that matters, for scientists, is reputation within the small community of peers regarded as possessing the appropriate qualifications to evaluate scientific work. For Merton the scientific ecosystem is a complex and fine-grained system of reputation management - and this reputational ecosystem has at its heart the act of publication.

Why is reputation so important in science? For Merton the answer to this question connects to the norm of 'communism' - in economists' terms, the status of science as a public good. It is in society's interests that scientific findings be made broadly accessible; but science is not intrinsically a public good - there is no necessity for scientists to share their knowledge in a way that allows that knowledge to be used by others. Indeed, there seem to be strong incentives for scientists *not* to do this. If knowledge can be 'enclosed' as intellectual property, then scientists can directly benefit from their research findings; by giving up intellectual property rights to those research findings,

scientists are giving up the principal means by which they can be rewarded for their work. What incentive structure could prompt scientists to engage in this apparently personally self-destructive - although socially beneficial - behaviour?

For Merton the reputational incentive structure of science is the answer to this question. Merton argues that, by publishing scientific findings, and making those findings available for broad public use, scientists give up intellectual property rights over their intellectual output in exchange for *reputational* property rights. Merton's work on science describes:

> the seeming paradox that in science, private property is established by having its substance freely given to others who might want to make use of it. (606)

Importantly, these reputational property rights are only granted with *priority* of publication. If a working scientist is the first to publish a research finding, that scientist gains the 'reputational property rights' to that finding. If she is the second, or the third, or the fourth to publish the finding, the reputational gains from the publication are meagre. Science is thus a 'first past the post' system, in which scientists compete in a race to acquire reputational property. For Merton this feature of the scientific incentive structure explains the frequency of 'multiples' - occasions where the same, or almost the same, scientific finding has been published almost simultaneously by multiple research teams: such teams are competing to be the first to publish a result that is clearly within reach, and thereby reap the 'winner takes all' rewards.

Dasgupta and David, in their 'new economics of science', take over this Mertonian explanatory scheme. They agree that science is a first past the post competition for reputational property, and that such property can be achieved by publishing findings in public venues where they are available for general use. For Dasgupta and David - as for Merton - this system has notable strengths and some notable weaknesses.

Among the strengths of the system is the fact that it incentivises broad dissemination of final research findings. Among the weaknesses is the fact that it encourages inefficient duplication of research work, as multiple research teams compete to be the first to publish the same finding. Still more importantly, for Dasgupta and David, the first past the post system disincentivises the sharing of the intermediate inputs of scientific research. If a

researcher shares their preliminary findings, or data, or computer code, etc., prior to the publication of their research results, their competitors might use those intermediate inputs to achieve priority of publication. In this scenario all reputational rewards accrue to the researcher's rivals, rather than to the researcher themselves. This system therefore incentivises the hoarding of intermediate inputs, because priority of final research publication is all that 'counts' in terms of the scientific reputational ecosystem.

Dasgupta and David share Merton's diagnosis of this inefficiency in the production of scientific knowledge. But Dasgupta and David build on this diagnosis, by formalising it using the tools of game theory.

2.4.2 Science as prisoner's dilemma

For Dasgupta and David, this Mertonian picture of scientists hoarding intermediate inputs can be formalised as a prisoner's dilemma game. As is well known, a prisoner's dilemma is a form of non-cooperative game involving two available strategies for each player: 'cooperate' or 'defect'. Both players would be better off if both players cooperate than if both defect. But 'cooperate' is also a strictly dominated strategy for each player: each player would be better off defecting than cooperating, no matter what the other player does. The matrix below shows the general form of the two-player prisoner's dilemma.

	C	D
C	x, x	z, w
D	w, z	y, y

Where $w > x > y > z$

For Dasgupta and David, this game usefully represents the problem identified by Merton around the hoarding of intermediate inputs.

Dasgupta and David imagine a scenario in which two research teams are working on the same problem. Solving the problem requires solving two subsidiary problems: a and b. Research team A has solved problem a and research team B has solved problem b. If both teams share their preliminary findings with the other, then both can immediately solve the overall problem, and move to the write-up stage of the research process. Which team achieves priority of discovery is then a matter of chance. If one team shares their preliminary findings, and the other team does not, then

the team who does not share their findings will publish first, and accrue all reputational gains. If neither team shares their preliminary findings, then both must continue to expend time and resources in further research. Clearly, here, the socially optimal outcome is for both teams to share their preliminary findings, but the Nash equilibrium is for both teams to refuse to share. This is a classic prisoner's dilemma; for Dasgupta and David this prisoner's dilemma encapsulates the social inefficiency of the priority reputational system.

As is well known, while the two-player prisoner's dilemma cannot be 'solved' in a one-round game, cooperative outcomes are possible in iterated versions of the game. In iterated versions of the game, conditionally cooperative strategies based on the other player's past play become available - for example the 'grim trigger' strategy, in which the player plays 'cooperate' as long as their opposite number also plays 'cooperate', but plays 'defect' for all time as soon as their opposite number plays 'defect' (Friedman 1971). Such conditionally cooperative strategies allow players to 'sanction' other players for noncooperative play, incentivising ongoing cooperation and making cooperative equilibria possible even in games that have prisoner's dilemma structures for every individual round. More strongly, as articulated by the 'folk theorem', any overall outcome in which both players' payoffs exceed those available from universal defection can be an equilibrium provided players' discounting is low enough (Fudenberg and Maskin 1986). Furthermore, such cooperative equilibria are possible even in finitely repeated versions of the game, provided the assumption of common knowledge is abandoned (Kreps and Wilson 1982).

As Dasgupta and David emphasise, however, scientific competition rarely produces a two-player version of the prisoner's dilemma. Rather, most real-world scientific researchers are participating in n-player iterated prisoner's dilemmas. I will postpone discussion of the literature on the n-player iterated prisoner's dilemma until later in the thesis. For now, it is enough to emphasise that, for Dasgupta and David, the problem of the efficiency of scientific production is centrally connected to the problem of achieving cooperative outcomes in an n-player prisoner's dilemma. I will return to this problem, and this model, in later chapters, when considering how different institutional structures of scientific production create more or less efficient knowledge production processes.

2.5 The economics of scientific knowledge

If Dasgupta and David's new economics of science helps reframe the problem of *efficiency* in relation to the specific institutional structure of science, the 'economics of scientific knowledge' aims to do the same for the questions of *credibility*. Dasgupta and David drew on the mid-twentieth century sociology of science developed by Robert K. Merton. The economics of scientific knowledge draws on the next wave of sociology of science: the 'science studies' tradition that broke with Mertonian functionalism, in the early- to mid-1970s, to develop a more 'radical' brand of 'constructivist' sociological explanation.

Merton's sociology of science, as I have already discussed, was characterised by attention to the norms and incentive structures associated with science as an institution. But although Mertonian sociology aimed to give sociological explanations for the professional behaviour of working scientists, it did not aspire to give sociological explanations for the *content* of scientific research itself. The question of the validity, rationality or truth of scientific findings was not, for the Mertonians, a problem to which sociologists could usefully contribute - except to the extent that the objects of scientific study were themselves social. Mertonian sociology of science thus divided the study of science into the question of social institutions (the province of sociologists) and the question of scientific truth (the province of scientists themselves - and, for some, philosophers of science).

In the later decades of the twentieth century, a new wave of sociologists of science challenged this understanding of the appropriate limits of sociological explanation. The founding classic of this new tradition was Thomas Kuhn's 1962 *The Structure of Scientific Revolutions*. Kuhn's account of the transition between scientific 'paradigms' - and of the role of background assumptions in structuring the 'normal science' that took place within those paradigms - was widely taken as a fundamental challenge to the narrative of linear scientific progress that informed both traditional philosophy of science and Mertonian sociology of science. Kuhn's suggestion that social factors with no obvious epistemic function might play a fundamental role in the formation of scientific consensus thus provided the initial philosophical basis for a subsequent wave of 'critical science studies'.[1]

[1]See Zammito (2004) for a detailed and insightful intellectual history of post-Kuhnian sociology of science.

Within sociology the first major tradition that picked up on and developed this Kuhnian approach was the 'strong programme' in the 'sociology of scientific knowledge'. In a series of publications in the early to mid 1970s - most prominently Bloor's 1976 *Knowledge and Social Imagery* - the strong programme theorists argued that Kuhn's insights demanded a fundamental restructuring of the way in which sociological research of science was carried out. For Bloor, earlier sociological theories of science:

> divide behaviour or belief into two types: right and wrong, true or false, rational and irrational. They then invoke causes to explain the negative side of the division. Causes explain error, limitation and deviation. The positive side of the evaluative divide is quite different. Here logic, rationality and truth appear to be their own explanation. Here causes do not need to be invoked. (Bloor 1976: 6)

By contrast, Bloor's 'strong programme' aimed to be 'symmetrical' in its account of the formation of scientific content: *both* truth *and* falsehood should be analysed using the same set of sociological resources (Bloor 1976: 5). At the same time, an emerging tradition of feminist and critical race theory aimed to analyse the way in which sexism, racism, and other apparently non-scientific social factors have a major but often unacknowledged impact on the formation of scientific content (e.g. Harding 1986, 1993; Haraway 1988; Jasanoff 2004; Mills 2007). Together with the extensive empirical work informed by the strong programme, these traditions formed an influential wave of 'critical science studies'.

This new, more 'constructivist' tradition in the sociology of science was widely criticised. Many such critics saw the 'sociology of scientific knowledge', not as extending the reach of the social scientific analysis of science, but as challenging the epistemological authority of scientific truth-claims. In particular, the strong programme's principle of 'symmetry' was widely attacked for undermining the claim to objectivity of scientific knowledge. If true and false scientific statements were to be treated 'symmetrically', critics asked, then how could the important differences between true and false statements be asserted? (Laudan 1981; Sokal and Bricmont 1999) Did the strong programme's principle of symmetry not commit the sociologist of scientific knowledge to an untenable epistemological relativism?

I discuss these debates in more detail in Law and Pepperell (2017). For the purposes of this chapter, I will limit my discussion to the ways in which

the sociology of scientific knowledge was picked up and interpreted by the emerging field of the *economics* of scientific knowledge.

This latter approach - the economics of scientific knowledge (ESK) - was first articulated as its own distinct research programme by Wade Hands in a series of papers from the mid 1990s (Hands 1994b,a, 1997). In his 1994 essay on the sociology of scientific knowledge, Hands defines the emerging subdiscipline as follows:

> the economics *of science* would be the application of economic theory, or ideas found in economic theory, to explaining the behaviour of scientists and/or the intellectual output of the scientific community... On the other hand, the economics of *scientific knowledge* (ESK) would involve economics in a philosophically more fundamental way... that is, economics would be involved fundamentally in the epistemological discourse regarding the nature of scientific knowledge. (Hands 1994b: 87-8)

In the remainder of this section I will look at these ESK attempts to model the determination of scientific content. I will schematise the tradition into two approaches. *First*, there are approaches that aim to 'bracket' the question of which scientific hypotheses are correct or incorrect, in order to model the social processes of consensus-formation in exclusively social terms. Here I will focus on work by the philosopher Jesus Zamora Bonilla, and - once again - the economist Paul David. *Second*, there are approaches that posit an objective state of affairs, as a standard against which scientific theories can be evaluated, and that then analyse social decision-mechanisms in terms of their ability to approximate this state of affairs in scientific hypotheses. Here I will focus on the Popperian philosophical tradition, with a particular emphasis on the work of Alvin Goldman. The former category I will refer to as *symmetrical* approaches to the economics of scientific knowledge; the latter as *evolutionary*.

2.5.1 Truth as consensus: symmetry in the ESK

One prominent approach within the sociology of scientific knowledge, exemplified by Harry Collins (Collins and Cox 1976; Collins 1983), was to 'bracket' the question of scientific validity, to focus on consensus-formation as an exclusively *social* process. A prominent advocate of this approach

within the economics of scientific knowledge is the philosopher of economics Jesus Zamora Bonilla. In his 1999 paper 'The elementary economics of scientific consensus' Bonilla proposes a very simple model of consensus-formation. This model is a valuable starting point for the discussion of ESK, as its basic features are common to many of the approaches I will discuss below.

Bonilla's model looks at the process of consensus-formation around a single scientific proposition. He assumes that every researcher in the relevant scientific field must make a decision about whether or not to accept or endorse the proposition. Further, he assumes that each scientist's decision on this matter is a function of *two* factors.

On the one hand, every scientist has an individual epistemic value that they place on the proposition: the degree of confidence the researcher has in the proposition based on 'private' evidence, such as personal research, reasoning, intuition, etc. On the other hand, each scientist also takes into account the views of other scientists in their community. This might be because scientists are reputation-conscious, and do not wish to be isolated among their peer group; or it might be because scientists' views are taken to be useful signals of the evidence those scientists have themselves personally collected and analysed. For Bonilla, scientists take other researchers' views into account in a very simple way: they are influenced by the proportion of other scientists who endorse the proposition.

For Bonilla, then, if a scientist places a high epistemic value on a proposition, they will reject it only when the overwhelming majority of the community also rejects it; likewise, if a scientist places a low epistemic value on a proposition, it would take an overwhelming community consensus in favour of the proposition to induce the scientist to accept it. Finally, if the epistemic value a scientist gives to a proposition is neither very high nor very low, their decision about whether or not to endorse the proposition may be very sensitive to other scientists' decisions.

Bonilla shows that this 'network effects' model of scientific decision-making is capable of producing multiple equilibria. Which scientific consensus emerges from this model is strongly dependent on starting-conditions - whether these are understood as existing scientific beliefs, or as existing expectations about future beliefs. In Bonilla's words:

> if 'scientific knowledge' is what a scientific community accepts as such, then the epistemic criteria of individual researchers are not enough to explain what has to be taken as 'scientific knowledge':

> it is also necessary to take into account the mutual interdependence between different individuals. (Bonilla 1999: 11)

Furthermore:

> even when this interdependence is considered, there can exist several 'social configurations of knowledge', and the actual one will normally be chosen by more or less accidental reasons. (11)

Bonilla takes his approach to offer support, within a rational-choice framework, for claims that have been prominently asserted - and debated - within the philosophy and sociology of science, specifically ideas about the underdetermination of theory by evidence (Quine 1951), and the corresponding role of social factors in the determination of scientific content (Kuhn 1962; Bloor 1976).

In later work, Bonilla and coauthors build upon this starting-point to develop more sophisticated models of scientific decision-making. Most notably, in a 2002 paper on a 'contractarian approach' to scientific inference, Bonilla introduces an important additional factor: the collective determination of a threshold epistemic value that scientific claims must pass before they can legitimately be accepted by scientific researchers.

This is not, however, the only direction in which the basic model can be extended. An important alternative approach was pioneered by Paul David's 1998 paper 'Communication norms and the collective cognitive performance of "invisible colleges"'. Working independently from Bonilla - his paper was published the year before Bonilla's 'Elementary economics' - David develops an almost identical basic model of scientific consensus-formation. For David, as for Bonilla, scientists' decisions about what hypotheses to endorse are shaped by two factors: individual preferences; and the 'reputational effects' of other scientists' views. However, where Bonilla aims to add sophistication to this model by introducing a contractarian apparatus around the determination of collective norms that constrain decision-making, David adds sophistication through two alternative strategies. First, he locates the process of scientific decision-making within a network of scientists - a set of 'invisible colleges' - the structure of which influences the outcome of the collective decision process. Second, he takes scientists to be concerned not just with their reputation among their colleagues, but with the possible divergence between their *short term* and *long term* reputation.

For David, as for Bonilla, a scientist considering whether to endorse a proposition is aiming to maximise the esteem in which she is held. Endorsing a proposition that is widely rejected by her peers will lead to low reputational outcomes; endorsing a proposition that is widely accepted will lead to high reputational outcomes. However, the scientist is also aware that group consensus may change. For many scientists, ultimate validation by the group after initial minority status is a *more* desirable outcome than being in agreement with the group throughout. And how scientists differentially weight these outcomes can have an important influence on what consensus is ultimately reached.

For David, moreover, this process of consensus-formation is best modelled not as the determination of a Nash equilibrium, but as as a Markov chain process within a graph-theoretic environment - a version of the 'voter model' (Clifford and Sudbury 1973). Each scientist's choice about whether or not to endorse a proposition is influenced by the behaviour of her neighbours on the graph. The system overall will ultimately 'lock in' to one of its 'extremal solutions'; that is to say, scientists will ultimately converge on a consensus around whether to accept or reject any given proposition. But the ultimate outcome, and the route taken to it, is both stochastic and path-dependent - in other words, 'history matters'. In David's model the outcome is influenced by a range of factors: the weighting scientists place on different reputational outcomes; the structure of the graph-theoretic environment; the 'openness' or 'closure' of the network to new entrants; and the willingness of researchers to share information with their peers. I will return to several of these factors, and to David's model itself, in chapter six's discussion of the epistemic behaviour of traditional, 'Mertonian' science.

Despite these differences, Bonilla's and David's models share the common ground that - although individual agents have epistemic preferences - the model itself is agnostic as to the value of these preferences. For some advocates of the strong programme in science studies, this feature of Bonilla and David's account of consensus-formation is a strength. For many theorists, however, an account of scientific consensus-formation that is not also capable of *evaluating* the consensus with reference to an objective external standard - or at least some commonly held normative criteria for selecting between competing scientific claims - is significantly flawed.

An alternative tradition within the economics of scientific knowledge aims to address this issue, by giving an *evolutionary* account of the development of scientific claims. Analysing the development of science by analogy with

biological evolutionary processes, this tradition understands the collective endeavour of empirical investigation as *selecting between* competing scientific hypotheses, in a way that will, over time, tend to eliminate the least empirically credible, and select the most empirically credible, of the hypotheses available. In the remainder of this subsection, I will look at a set of theorists who have aimed to develop an account of science within this tradition.

2.5.2 Evolutionary social epistemology

The foundational classic within the evolutionary tradition is Karl Popper's *The Logic of Scientific Discovery* (Popper 1959). For Popper, science can best be understood as a process of 'conjectures and refutations' (Popper 1963). Scientific researchers propose a wide range of hypotheses to explain natural phenomena; experiments and observations are then carried out to test these hypotheses, and the hypotheses that are incompatible with the evidence are rejected. In this way a field of possible claims is narrowed down by an empirical selection process. For Popper, this account addresses many classic problems in the philosophy of science - most notably the problem of induction - without undermining progressive accounts of scientific discovery. It further provides a criterion - falsifiability - that allows scientific claims, which are vulnerable to empirical refutation, to be clearly differentiated from non- or pseudo-scientific claims, which are not.

This Popperian picture of scientific research has some formal similarities with the account of evolution by natural selection. There is a diversity of theories; selection among those theories by environmental factors; and mutation of existing theories as scientists innovate on the basis of past knowledge. In Popper's words:

> the growth of our knowledge is the result of a process closely resembling what Darwin called "natural selection"; that is, the natural selection of hypotheses (Popper 1979: 261)

This idea was adopted and developed by a subsequent generation of 'evolutionary epistemologists' - often generalising the Popperian picture of scientific progress to knowledge-creation processes in general. Donald Campbell - one of the major figures in the subsequent development of evolutionary epistemology - puts it as follows:

> A blind-variation-and-selective-retention process is fundamental to all inductive achievements, to all increases in knowledge, to all increases in fit of system to environment. (Campbell 1960: 380)

For most twentieth century figures in this evolutionary tradition, however, (e.g. Lakatos and others 1970; Hull 1988; Wible 1998), the parallel between Darwinian natural selection and the process of scientific discovery is primarily qualitative. Thus Wible, for example, while using microeconomic tools to model the cost-benefit decisions of scientists choosing whether or not to share data, restricts himself to qualitative remarks in his development of "an evolutionary conception of rationality in science" (1998: 193-206)

One important late twentieth century contribution to the project of greater modelling specificity is Philip Kitcher's 1993 *The Advancement of Science*. Kitcher provides detailed microeconomic models of the social selection processes he analyses. And unlike many Popperian philosophers of science, Kitcher engages deeply with the Kuhnian ideas of the critical science studies tradition.

Kitcher's core idea is that each scientist can only personally evaluate a very tiny fraction of the claims they are expected to endorse. The great majority of any working scientist's scientific worldview must therefore be derived, in one way or another, from authority. For Kitcher, however, unlike Bonilla and David, there are *three* core forms of scientific authority:

> First, there is the general epistemic dependence on the past that figures in everyone's early intellectual ontogeny... Second, at the time of entry into the scientific community, novices endorse a communitywide conception of legitimate epistemic authority... Third, during the course of individual research, scientists interact with one another, adopting the claims made by *some* of their colleagues, investigating the proposals of others, ignoring the suggestions of yet others... (1993: 306)

The focus of Kitcher's work is this final form of authority - the authority vested by members of the scientific community in the claims of their colleagues.

This emphasis on the authority of others' claims (cf. Hardwig 1985) has been developed by others working in the emerging field of social epistemology. Notably, Alvin Goldman's *Knowledge in a Social World* proposes a

research programme that Goldman refers to as 'veritistic social epistemology'. (Goldman 1999) Practitioners of veritistic social epistemology assume that our knowledge formation practices are intrinsically social: we rely on the authority of others for many of our beliefs, and the conclusions we draw from our own direct observations of the natural world are influenced by a socially formed interpretive framework. The social practices involved in belief formation are therefore crucial to the 'veritistic outputs' of our institutions.

For Goldman, different social practices and social structures are likely to lead to different 'veritistic outputs'. In his words:

> veritistic social epistemology aims to evaluate social practices in terms of their veritistic outputs, where veritistic outputs include states like knowledge, error and ignorance. What I am contemplating, then, is an evaluative structure of two parts. States like knowledge, error, and ignorance have *fundamental* veritistic value or disvalue. Practices have *instrumental* veritistic value insofar as they promote or impede the acquisition of fundamental veritistic value. (Goldman 1999: 87)

For Goldman, then, the task of veritistic social epistemology is to compare the knowledge outputs of different social practices and institutions. Moreover, Goldman suggests that the tools of economic modelling are well suited to this task. In Goldman and Shaked (1991), for example, the authors propose a model of scientific practice in which scientists seek to maximise expected utility, "determined solely by professional success" (1991: 31). Goldman and Shaked show that a community of Bayesian agents motivated by scientific reputational outcomes, rather than directly oriented to truth, will under reasonable assumptions still generate positive knowledge outcomes. More broadly, their approach serves as another model for the application of formal economic modelling resources to epistemological questions. In the next chapter I will explore in more detail some of the other modelling resources available within this framework.

2.6 Agent-based modelling

The previous two sections have looked at the microeconomic analysis of scientific institutions and the dynamics of scientific knowledge. In this final section of the chapter, I look at an analytic approach that is increasingly

used to analyse both of these dimensions of scientific practice, and that I will be drawing on extensively in later chapters of the thesis: agent-based modelling.

Agent-based modelling (ABM) has only relatively recently emerged as a significant component of the economist's modelling toolkit. The concept of cellular automata - systems in which discrete cells change their states based on the states of the cells in their 'neighbourhood' - has existed since the 1940s, when it was developed by John von Neumann and Stanislaw Ulam (Burks and Von Neumann 1966).[2] In 1969, Schelling introduced a simple agent-based model to the mainstream of economics, with an influential paper that used an ABM to analyse possible dynamics of segregation (Schelling 1969). In 1996, Epstein and Axtell expanded the scope of agent-based modelling in the social sciences, arguing that 'growing' social systems 'from the bottom up' could be seen as a criterion of social-scientific explanatory adequacy, and illustrating this suggestion with a model of multiple different categories of agent interacting in an environment they called 'sugarscape' (Epstein and Axtell 1996). However, it is only with the more recent growth in computational economics and computer science that agent-based modelling approaches have been more widely adopted within the discipline of economics (Tesfatsion and Judd 2006). Increasing numbers of economists now argue that agent-based modelling fills an important analytic niche within the discipline (Farmer and Foley 2009; Ballot, Mandel, and Vignes 2015; Hamill and Gilbert 2016).

Like all kinds of formal modelling - and as I discuss in more detail in the next chapter - agent-based modelling is a framework that suits some analytic tasks and contexts better than others. Its advantages over more traditional formal modelling approaches include the following.[3] First, because agent-based modelling studies the interaction of many simple components to create potentially complex patterns, it is particularly well-suited to the analysis of 'emergent' phenomena - that is to say, aggregate behaviours of a system that cannot be understood via the analysis of any specific component part of the system.[4] Second, agent-based modelling is well-suited to the study of the

[2]For an overview of the history of cellular automata and agent-based models in general, see Wilensky and Rand (2015: 431-446).

[3]For a discussion of the methodological strengths of agent-based modelling, see Tesfatsion (2006).

[4]There is a debate within the philosophy of science about the relative merits of 'strong' versus 'weak' understandings of emergence (Chalmers 2006). This thesis is interested throughout in a 'weak' sense of emergence: the emergent properties of the scientific systems

behaviour of diverse categories of social actor: different kinds of social actors behaving in different ways. Third, and finally, agent-based modelling is a valuable tool for the study of *evolutionary* dynamics, focused on selection among variation. For this reason, agent-based modelling tools align particularly well with the kinds of evolutionary social epistemology that I discussed in the previous section. As I will discuss in more detail in Part II, in this thesis I analyse science as a complex evolutionary system generating emergent properties. Agent-based modelling will therefore prove to be a particularly useful analytic tool in studying the kinds of scientific dynamics I discuss in later chapters.

In the remainder of this section, I will look at some of the most important applications of agent-based modelling to the study of science dynamics.

2.6.1 Agent-based models of science

In earlier sections of this chapter, I have discussed a range of formal models of science as an institution, including those of Dasgupta and David (1994), Kitcher (1993), and others. More recent scholars have argued that the kinds of idealisation associated with traditional economic modelling have limited our understanding of the dynamics of scientific practice. In particular, Muldoon and Weisberg (2011) have made the case that traditional marginalist and rational actor approaches assume unrealistically high degrees of knowledge by social actors, both of other scientists' actions and of the likely rewards associated with any given research strategy. Muldoon and Weisberg argue that if assumptions such as these are relaxed or abandoned, our picture of the dynamics of scientific practice changes considerably. Moreover, agent-based modelling - with its ability easily to capture the effects of bounded rationality, local knowledge, and heterogeneous agents - is particularly well-suited to addressing these shortcomings of the traditional economics of science and scientific knowledge.

In this spirit, there is a growing set of related research programmes in the agent-based modelling of science, focused on both the behavioural and the epistemic dimensions of scientific practice. Some researchers have sought to use agent-based models to 'grow' from the 'ground up' behaviours that match the patterns of publication and citation empirically observed in studies

I study here are not assumed to have any 'ontological' or causal status independent of the analysis of the component parts of that system.

of scientific communities. Thus for example Sun and Naveh (2009) build on Gilbert's (1997) equation-based simulation of a research mechanism that can generate the empirical pattern of authorship credits observed by Lotka (1926). For Sun and Naveh, modelling the cognitive and strategic decisions of agents authoring scientific papers allows researchers to 'experiment' with the impact of different cognitive strategies on research outputs.

Much recent work in agent-based modelling of science has focused, however, on the epistemic dimensions of science as a social system. One productive strand of research here is focused on allocation of scientific research attention, and the impacts of the division of cognitive labour on scientific outcomes. Here the most influential figure is Zollman, who in a series of papers (Zollman 2007, 2009, 2010; Kummerfeld and Zollman 2015) has developed a set of models of scientific research dynamics which allow researchers to investigate the epistemic impact of different approaches to the division of cognitive labour. Some of Zollman's findings are apparently counter-intuitive. For example, in a 2007 paper Zollman shows that in some scenarios scientists sharing *more* information can result in *poorer* epistemic outcomes (Zollman 2007). The reason for this surprising finding is that high levels of information sharing can produce rapid consensus-formation, including in scenarios where a small number of early misleading findings can result in rapid consensus around a faulty conclusion or method. In Zollman's words:

> this model suggests that in some circumstances there is an unintended benefit from scientists being uninformed about experimental results in their field.... [W]hen we want accuracy above all else, we should prefer communities made up of more isolated individuals. (Zollman 2007: 586)

At the same time, Zollman also emphasises that too great or too persistent diversity of scientific opinions or approaches can result in a scientific community that never approaches consensus, and that therefore cannot fulfil its function as a producer of reliable knowledge. In this sense, Zollman argues, the goal of epistemic systems should be *transient* diversity (2009). Initial diversity can be achieved either by limiting the degree of contact between scientists, or by giving agents extreme priors. However, "[i]f agents are both extreme in their initial beliefs and limited in their access to information, the initial diversity is never abandoned" (Zollman 2009: 19). The challenge for science as an epistemic system, for Zollman, therefore lies in maintaining

sufficient cognitive diversity to achieve reliable ultimate consensus, without preventing a consensus from ultimately emerging.

Similar approaches to this problem space have been taken by a range of other researchers. De Langhe (2014) constructs an agent-based model that aims to capture the relationship between scientists exploring existing theories and developing new ones. De Langhe uses this model to analyse the optimal division of cognitive labour between consensus and innovation in theory creation and exploration. Borg et al. (2018) argue that modelling approaches that emphasise only theory creation and information exchange are too idealised, and propose an 'abstract argumentation' agent-based model, building on the work of Dung (1995), which captures the role of argumentation in scientific discursive systems. This model suggests, against Zollman, that "increased information sharing tends to be epistemically beneficial" (Borg et al. 2018: 305).

At the same time, a parallel research program has focused on the metaphor of the 'epistemic landscape' as a tool for modelling science dynamics. Here the key paper is Weisberg and Muldoon's 2009 'Epistemic landscapes and the division of cognitive labor'. Weisberg and Muldoon propose a three-dimensional 'epistemic landscape', in which the two horizontal dimensions represent a research approach or topic, and the third dimension of height represents the epistemic significance of a research approach. Scientists are agents which can move around this epistemic landscape - much as agents in Epstein and Axtell's classic 'sugarscape' model can move around their own simulated landscape, searching for and consuming resources (Epstein and Axtell 1996). These agents seek the most significant research findings - i.e. the highest locations on the landscape - but they use different search strategies to do so. Weisberg and Muldoon use this modelling approach to investigate the question of which exploratory strategies are likely to be most successful. They find that research communities that include at least some 'mavericks' - researchers who try to avoid research approaches pursued by colleagues - are much more epistemically successful than others.

This 'epistemic landscape' approach to modelling the epistemic dimensions of scientific research has also been adopted by a range of other researchers (e.g. Alexander, Himmelreich, and Thompson 2015; Thoma 2015; Pöyhönen 2017). Grim (2009) similarly has agents exploring an epistemic landscape, but where Weisberg and Muldoons' agents observe the actions of their neighbours on that landscape, Grim's agents are also part of a social network. Grim's agents can therefore learn from other members of their so-

cial network about quite distant features of the epistemic landscape. Grim's conclusion is that, above a certain small level of interconnectedness, larger and more connected social networks *decrease* epistemic performance, because they lead to rapid 'clustering' around early successes, even where those successes may be relatively poor locations in the epistemic landscape as a whole - the same finding as Zollman (2007). Thoma (2015) modifies the behavioural rules of Weisberg and Muldoon's agents to better capture scientists' ability to choose new research areas that may differ significantly from their current work, by allowing agents to move to research approaches outside their immediate 'neighbourhood'. Thoma concludes that, with this and other modifications, the epistemic landscape model supports the intuitive conclusion that diversity of research strategies can have overall epistemic benefits. Pöyhönen (2017) argues that, once Weisberg and Muldoon's relatively smooth epistemic landscape is replaced with a 'rugged' landscape including many smaller 'unproductive' peaks, social learning with a diverse array of epistemic strategies produces the best overall epistemic outcomes - that "no learning strategy is *per se* more rational than others" (Pöyhönen 2017: 4536).

These models are typically 'high level' and abstract, illustrating conceptual possibilities rather than attempting to model specific empirical dynamics. Increasingly, however, agent-based modellers have also sought to calibrate their models using empirical data. Thus, for example, Harnagel (2018) modifies Avin's (2017) agent-based model of science funding to calibrate the core features of the model using empirical data on scientific publications drawn from the Thompson Reuters Web of Science database. In this way, Harnagel constructs an 'epistemic landscape' based on real scientific community judgements of scientific significance, as proxied by citations. Harnagel argues that this kind of 'mid-level' model, which combines 'top down' and 'bottom up' methods of model construction, provides a better guide to policymakers than purely theoretical approaches.

Where the models discussed above have primarily focussed on the impact of different agents' research strategies on epistemic outcomes, a growing literature also foregrounds the role of *institutional* constraints on the behaviour of the epistemic system. Thus for example Chavalarias (2017) constructs an agent-based model of a 'Nobel game' in which scientists compete to publish and refute findings. Chavalarias argues that this model highlights a trade-off between the speed and the reliability of the scientific research process, and demonstrates the epistemic downsides of a 'publish or perish' institutional model. Sobkowicz (2015) likewise studies the ways in which funding avail-

ability influences the evolution of scientific research interests, and suggests that greater transparency in the funding allocation process may reduce the power of research 'cliques' to direct funding to their own members.

A number of researchers have recently used agent-based models to study the institutional structure of the peer review process. Bianchi and Squazzoni (2015) look at the impact of the number of reviewers on the quality and efficiency of the peer review process, under different scenarios of reviewer reliability. They use their agent-based model to make the case that multiple reviewers typically improve the reliability of the epistemic system, but that these reliable outcomes may come at an unsustainably high cost in terms of the allocation of cognitive resources. Kovanis et al. (2017) model the scientific publishing process in order to evaluate the efficiency and dissemination impacts of five different peer review systems. They argue that 'review sharing' systems, in which rejected manuscripts are resubmitted to new journals accompanied by their previous peer reviews, are preferable to both the existing peer review system and more 'open' approaches in which manuscripts are immediately published on the open web. Righi and Takács (2017) study peer review as a social dilemma, in which the default dominant strategy for both reviewers and paper authors is to produce low quality work. The authors use their model to show that reputational dividends associated with publication in high quality journals may incentivise high quality scientific work, but their model is unable to make high quality *reviewing* a dominant strategy; the authors conclude that "it is not easy to find the way out of this puzzle" (Righi and Takács 2017: 604).

One implication of focussing on the institutional dimension of the scientific system is that 'evolution' of scientific behaviour can be studied without any assumption of strategic or even intentional behaviour on the part of individual scientists. Thus, for example, in a recent and important paper Smaldino and McElreath (2016) model what they call 'the natural selection of bad science'. Where Gilbert and others studied the 'descent with modification' of research *content*, Smaldino and McElreath model the 'descent with modification' of research *methods*. Specifically, Smaldino and McElreath look at the differential success of methods of differing reliability. The authors argue that unreliable research methods that nevertheless produce striking and publishable findings are likely to spread through the research community, because of their output's appeal to hiring and funding bodies. By contrast, reliable methods that do not produce such striking findings are likely to be selected *against*. This dynamic, Smaldino and McElreath argue,

results in the 'natural selection of bad science', a conclusion which connects to broader worries about the 'replicability crisis' in science, which I will discuss in chapter seven. This finding also poses an important challenge to the Popperian tradition of evolutionary social epistemology, for it implies that the selection pressures of our existing scientific institutions may be selecting *against* truth and reliability in scientific findings and methods. If this vision of the scientific research process is correct, science can indeed be seen as an evolutionary process, but it is not necessarily a reliably truth-oriented one.

In a follow-up paper, which is particularly relevant to the themes of this thesis, Smaldino, Turner and Contreras (2019) look at the consequences for this 'natural selection of bad science' of a set of scientific institutional changes: allocation of science funding by unconventional mechanisms like lotteries; the publication of negative results; and the improvement of the reliability of the peer review process. Smaldino et al. refer to these latter two institutional changes as 'open science' - and they conclude that a shift towards 'open science' institutional structures can have a significant impact on the reliability of scientific findings. This result resonates - and is compatible - with the argument of this thesis. However, the institutional changes that Smaldino et al. discuss under the heading of 'open science' are *not* the institutional proposals on which this thesis focuses. Where Smaldino et al. are focused on the impacts of improved peer review mechanisms and the publication of negative results, this thesis is focused on open *inputs* science: the broad public sharing of the intermediate inputs of scientific production. In Parts II and III of the thesis I will explore the impact of open inputs science, in the sense in which I defined this term in chapter one, on some of the epistemic dimensions of the scientific production process that the literature discussed in this section has tackled.

2.7 Summary and conclusion

In this chapter I have given an overview of some key elements of the economics of science.

The 'old' economics of science, I suggested, is typically associated with the mid-twentieth-century neoclassical revolution. This old economics of science saw science as a public good, best supplied by government subsidies and investment, funded by taxation. This picture integrated - albeit in an ad hoc manner - with the neoclassical growth theory associated with Solow: in the

Solow model technological development is the ultimate driver of continuing growth. For the neoclassical approach, technological development is itself derivative of scientific research, in turn derivative of state investment.

By contrast, in the modern economics of science, this neoclassical picture is challenged along several dimensions.

First, in growth theory, the Solow model has been partially supplanted by a range of endogenous growth theories. Second, the 'new economics of science' has aimed to analyse the micro structure of specifically scientific institutions, by producing formal models of the incentive structure of scientific knowledge workers. Third, a range of theorists in the 'economics of scientific knowledge' have used the tools of economic theory to analyse the epistemic dimensions of scientific production. Finally, agent-based modelling tools have facilitated the formal modelling of more evolutionary approaches to the study of science dynamics.

In later chapters of the thesis I will draw on these approaches to develop a set of original models of science as an institution. In the next chapter, I discuss the methodological challenges posed by that project, and in particular by modelling science as a knowledge institution.

Chapter 3

Modelling institutions

3.1 Introduction

Chapter two surveyed the economics of science, and identified the two elements of scientific practice that will be the focus of the thesis: efficiency of production, and credibility of output. In this chapter I will discuss the analytic strategies I intend to use to study these issues.

To foreshadow, there are two methodological 'pillars' of the analysis I will carry out in this thesis. In the first place I will argue, following Douglass North and others, that we cannot understand scientific institutions without understanding the historically and socially contingent 'rules of the game' that structure them. This analysis of socially contingent rules of the game must be, I will argue, in the first place *qualitative*. We must look at the practices of scientists and other social actors, to see how institutions are constructed in social practice.

In the second place I will argue that, once we have described the rules of the game that structure a specific economic institution, we are then in a position to analyse the behaviour of actors within that institution using the toolkit of formal economic modelling. In particular, as I discussed in section 2.6.1, recent applications of agent-based computational modelling to the social study of science provide a rich framework for formal analysis of scientific dynamics.

In this chapter I outline the key methodological and theoretical resources I need to realise this modelling approach in the remainder of the thesis. In section 3.2, I survey the literature of institutional economics, to look at

a range of ways in which economists have studied the 'rules of the game' of economic life. I discuss institutional economics, 'old' and 'new', with a particular focus on the influence of Ronald Coase.

In section 3.3, I look in more depth at the relationship between the two different subtraditions within new institutional economics: those that emphasise transaction costs, and those that emphasise the rules of the economic game. I then focus on Elinor Ostrom's 'Institutional Analysis and Development' framework. This framework was initially developed to analyse common pool resource governance structures. In recent years, however, a range of scholars have extended Ostrom's framework to the analysis of scientific institutions. It is this qualitative analytic framework that I will adopt in later chapters of the thesis.

In section 3.4, I then briefly return to the recent history of formal models of scientific institutions, looking at the modelling of science using the tools of game theory and opinion dynamics modelling. In section 3.5, I discuss the problem of lack of realism in formal economic modelling. For many critics of formal models, the high abstraction of formal economic models renders them dubious representations of complex social reality. I survey the debate about realism and unrealism in modelling, and argue that the models put forward in this thesis meet Gibbard and Varian's methodological warrant for what I call 'caricatural' models.

Finally, in section 3.6, I discuss some epistemological concerns that have been raised by critics of contemporary science studies, in relation to philosophical debates over 'realism' and 'constructivism'. The thesis, I suggest, adopts a 'weak methodological realism', which is compatible with a range of metatheoretical or philosophical positions.

In sum, where chapters one and two outlined the *problem space* of the thesis, this chapter outlines the *methodological* and *metatheoretical* toolkit the thesis draws on to analyse that problem space. Taken all together, these first three chapters of the thesis thus assemble the resources that are required for the analysis that follows.

3.2 Institutional economics

I begin by reviewing the history of institutional economics, the broad tradition within which this thesis can be located. As is common in the literature (Hutchison 1984; Hodgson 1993; Rutherford 1996), I will here periodise in-

stitutional economics into 'old' and 'new' institutional economics, with a particular focus on the work of Ronald Coase as a bridge between the two periods. I then discuss the specific subtradition within institution economics that I will make use of in this thesis: Elinor Ostrom's Institutional Analysis and Development framework.

3.2.1 Old institutional economics

Contemporary institutional economics typically locates itself within a tradition that originated in the late 19th and early 20th centuries. This 'old' institutional economics was - with Marshallian value theory - one of the two central research paradigms in early- to mid-twentieth century US economics.[1] This tradition of 'old' institutional economics was quite varied, encompassing, for example, Veblen's evolutionary economics and instinct psychology (Veblen 1898, 1899, 1904); Commons' economic history, analysis of legal institutions, and studies of collective action (Commons 1934, 1932); Mitchell's quantitative historical studies of the business cycle (Mitchell 1913, 1927, 1959); and a range of other approaches (Stewart et al. 1919; Clark 1939; Ayres 1946).

Although these theoretical and methodological approaches were quite varied, they also shared some common features, well articulated in W.H. Hamilton's programmatic article on 'The institutional approach to economic theory' (Hamilton 1919). In that article, Hamilton argues that traditional Marshallian value theory presupposes an underlying and historically contingent institutional framework which it cannot itself theorise. For this reason, Hamilton argues, only an institutionalist perspective can be properly theoretical, in the sense of providing a general account of economic behaviour. At the same time, this generality of institutional economics is underwritten by the research programme's recognition of the historical contingency of any given institutional structure. A 'general theory' in one sense, institutional economics therefore typically rejects the possibility of 'general theory' in another. From Hamilton's perspective, the early 20th century debate between institutionalist and neoclassical economics can be seen as a debate not just between rival theoretical approaches, but also between historicist and anti-historicist philosophical frameworks.[2]

[1] For recent monographs providing historical overviews of old institutional economics, see Rutherford 2000, and Hodgson 2004.

[2] I use 'historicism' here, and throughout the thesis, to denote a metatheoretical orien-

In the middle decades of the twentieth century, the institutionalist school went into decline. Scholars have proposed a range of different causes of this decline: the failure of institutionalism to respond adequately to the experience of the Great Depression; the rise of Keynesianism; the development of the more sophisticated mathematical research programs of general equilibrium and econometric analysis; the decline of the Progressivist political project with which many institutionalists were associated; the social scientific movement away from evolutionary approaches; the scientific discrediting and political repudiation of eugenicism, which many early institutional economists endorsed; and other factors. From the middle decades of the twentieth century, however, economics took a newly mathematical neoclassical orientation, leaving little room for the more qualitative and historicist orientation of the institutionalists.

3.2.2 Coase

One important exception to this mathematical turn was Ronald Coase. Coase did not regard himself as an institutionalist (Coase 1998: 72), but he would become a 'founding father' for a later generation of new institutionalist economists. The Coasian contribution to institutional economics can be seen as grounded principally in two papers - 1937's 'The nature of the firm', and 1960's 'The problem of social cost'.

In 'The Nature of the Firm' Coase foregrounds the concept of transaction costs as a major influence on the scope and structure of economic institutions. Coase distinguishes between two ways in which production is coordinated: "a series of exchange transactions" on the market, or the decisions of an "entrepreneur coordinator, who directs production" within the firm (Coase 1937: 388). For Coase, this distinction implies a question:

> why co-ordination is the work of the price mechanism in one case and of the entrepreneur in another. (389)

tation that emphasises social and historical contingency. This usage differs from that of Popper (2002) and some others, who use 'historicism' to denote a metatheoretical orientation that seeks to uncover universal historical and social laws. The forms of 'historicism' criticised by Popper in his *The Poverty of Historicism* are therefore *anti*-historicist positions in the usage adopted here. My position differs from Popper's in vocabulary but not in substance.

To answer this question, Coase looks the differential costs associated with hierarchical planning and market exchange, in different economic contexts (392). For Coase, a firm is likely to include a category of transaction within its internal organisational structure if the costs of doing so are lower than the costs of the same transaction taking place within the market, and vice versa. This idea allows Coase to suggest an explanation for the size of any given firm:

> A firm will tend to expand until the costs of organising an extra transaction within the firm become equal to the costs of carrying out the same transaction by means of an exchange on the open market or the costs of organising in another firm. (395)

Coase's paper therefore provides a simple theoretical approach for explaining the existence and size of firms within a market economy: the existence, size and scope of any given firm can be explained by entrepreneurs' efforts to minimise transaction costs.

This use of the concept of transaction costs would become central to later institutional economics.[3] In 1967, Coase's 'The problem of social cost' proposed two more concepts that have been similarly influential.

In 'The problem of social cost' Coase argues that conventional Pigouvian welfare economics is 'asymmetrical' in its analysis of the negative externalities which state regulation or taxation can potentially resolve. Traditional welfare economics, Coase argues, looks only at the negative impact of a firm or industry on other social actors, without adequately considering the negative impact that regulation or taxation may have on the social benefits produced by that firm or industry (Coase 1960: 44). In making this case, Coase considers a set of scenarios in which:

- The only significant benefits and costs to be considered are economic;
- There are clearly defined property rights, agreed by all parties, associated with the relevant economic resources; and
- There are minimal transaction costs associated with economic bargaining and exchange.

[3]Though cf. Commons (1931) for a similar, earlier treatment of the transaction as the fundamental unit of economic analysis.

In scenarios with these three features, Coase argues, economic parties will naturally negotiate a mutually beneficial arrangement which maximises social output, obviating the need for government intervention - an idea that has become known as the 'Coase Theorem' (Cooter 1989).

As Coase stresses, in real-world situations the preconditions of the validity of the Coase Theorem rarely hold. Nevertheless, the 'theorem' has been influential in the subsequent literature, in part because it emphasises a theoretical possibility that had previously often been neglected, but also because it further illustrates the importance of the specification of property rights, and the role of transaction costs, in the outcome of economic bargaining.

'The Problem of Social Cost' contains one more point that has been important for subsequent work in institutional economics. In discussing the importance of correctly theorising property rights, Coase also discusses "a faulty concept of a factor of production" (Coase 1960: 43). As Coase puts it:

> We may speak of a person owning land and using it as a factor of production but what the land-owner in fact possesses is the right to carry out a circumscribed list of actions. (44)

For Coase, in other words, a factor of production is best understood, not as a physical resource, or even as an immaterial resource, but rather as a bundle of rights and obligations held by the relevant economic actors (cf. Sen 1981). Possession of an area of land is best understood as the generally socially recognised right to perform certain actions in relation to the land, and as an associated set of obligations for other social actors to refrain from performing those actions. These rights and obligations must be both socially recognised and socially enforced - paradigmatically by agents of state law enforcement, but potentially by other social actors too.

Thinking about factors of production in this way, Coase argues, increases the 'symmetry' of our analysis. But it has another important consequence: it redirects economists' attention from resources to institutions - from material factors of production to issues of law and social convention. These concluding remarks of Coase's article therefore parallel in an important respect Hamilton's proposal to embed value theory within an institutional economics, and Commons' focus on the importance of collectively established legal institutions: they suggest that while economists can study value or resources, they must understand that many of the key attributes of the objects of their study are *institutionally endowed*.

3.2.3 New Institutional Economics

From the mid-1970s, a new generation of economists began picking up these ideas, and developing them into a tradition that would eventually become known as the 'new institutional economics' (Klein 2000; Williamson 2000; Furubotn and Richter 2005). Although, again, 'new institutional economics' was quite diverse, the tradition can usefully be divided into two strands.

On the one hand, a tradition of 'transaction cost economics' emerged, building on Commons' concept of the transaction as the central economic category, and Coase's idea that the boundary of the firm is typically drawn in such a way as to minimise transaction costs. Associated particularly with the work of Oliver Williamson (Williamson 1975, 1979, 1985), transaction cost economics sought to 'operationalise' the insights of 'the nature of the firm' to provide a set of tools for the analysis of industrial organisation. I will look at this approach in more detail in section 3.3 below.

On the other hand, a range of alternative approaches to institutional economics were developing, which understood institutions as - in Douglass North's words - the "rules of the game" of economic life (North 1990: 3). For these institutional theorists, rules or conventions - institutions - constrain our economic actions, but these rules and conventions are, potentially, quite diverse. Although economists within this subtradition made use of the concept of transaction costs as a central analytic category, they often rejected the Williamsonian dichotomy between markets and hierarchies, and emphasised instead the variety within and beyond different kinds of institution.

'Rules of the game' theorists can in turn broadly be classified into two sub-groups. On the one hand are a set of scholars who create formal models of economic institutions - primarily rational choice and game-theoretic models. These scholars often understand institutions as a 'game' in the game-theoretic sense: a set of social constraints that influence economic actors' rational, strategic decision-making (e.g. Greif 1998, 1998; Aoki 2001).

On the other hand, a range of scholars have pursued a more qualitative approach to 'rules of the game' institutional analysis. Within economics, this approach has been particularly emphasised by Elinor Ostrom (Ostrom 1990, 2005). Although Ostrom draws on game-theoretic tools, her 'Institutional Analysis and Development' approach has also placed qualitative ethnographic case study analysis at the centre of its effort to understand both the formation and the stability of institutional norms and practices.

I will discuss Ostrom's Institutional Analysis and Development approach

in more detail in later sections. In the next section I look in more depth at the connection, and tensions, between the 'rules of the game' and the transactions costs frameworks within institutional economics.

3.3 The analytic status of the 'rules of the game'

One way to understand the basic 'transactions cost' approach is as an analytic framework that prioritises a specific form of causal explanation in its accounts of economic phenomena. For transaction cost economics, transaction costs are, all else equal, taken to be both analytically fundamental and causally prior to the 'rules of the economic game'. Shifts in economic governance structures - the 'rules of the game' that structure economic life - can, from the basic transaction costs perspective, be explained in terms of shifts in the costs of the underlying transactions those rules govern.

In this sense, transaction cost economics can be placed within a larger category of institutional political economy - encompassing Ayres' old institutionalist account of economic progress (Ayres 1946); traditional Marxist 'base-superstructure' accounts of economic change (Cohen 2000); and some other forms of 'economic determinism'. For all these categories of economic explanation, changes in governance structures are driven by changes in the underlying economic phenomena being governed. Moreover, for these perspectives it is the task of those instituting governance structures to select the governance structure that best corresponds to the underlying economic phenomena. Although there are more sophisticated versions of the transaction costs perspective - as I will discuss in a moment - its explanatory foundation is thus a simple claim about a dominant causal relationship running *from* transaction costs *to* the 'rules of the game'.



This is one 'Coasean' picture of the relationship between transaction costs and the rules of the game. However, as I discussed in the previous section, it is not the only perspective on this issue that can be derived from Coase's work. In the concluding paragraphs of 'The problem of social cost', Coase proposes a perspective which, if elaborated consistently, implies a very different explanatory framework.

From this analytic perspective - what I am calling 'rules of the game' Coaseanism - a economic good is not simply the physical or immaterial good, but rather or also the right or entitlement to make use of a physical or immaterial resource in a particular way.

Moreover, the economic actors that possess these rights can themselves, analogously, be understood as instituted by legal principles that establish legal entities capable of *possessing* rights.[4] Thus a person may (or may not) be legally recognised as capable of claiming ownership of a good, and so may a corporation or a component of a state or transnational body. From this perspective, the 'rules of the game' typically *precede* transactions, both causally and analytically, because the social relations of property ownership and exchange norms which are preconditions of transactions are themselves instituted by the rules of the economic game.

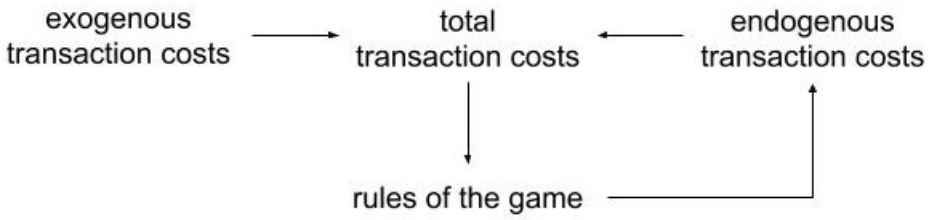
This perspective, which sees factors of production and, by analogy, those who possess them, in terms of socially-instituted rights and entitlements, is no less Coasean than the basic transaction costs approach - yet this Coasean insight has very different analytic implications from those derived from 'The nature of the firm'. This Coasean perspective explains transaction costs *in terms of* the rules of the economic game.

rules of the game ⟶ transaction costs

We thus have two quite difference categories of causal explanation and analytic framework associated with the new institutional economics, both derivable from different dimensions of Coase's work.

In practice, the tradition of New Institutional Economics draws on both of these 'ideal typical' subtraditions. One example of this is the work of Oliver Williamson. Although Williamson is typically, and rightly, understood as the major figure in the contemporary transactions cost approach, his work is clear that transactions and the rules of the game that govern them are in practice often mutually constituted. For example, one of the key elements of Williamson's argument is the transformation of the relationship between economic actors that takes place once a contract has been established between them (see e.g. Williamson 1985: 62; Riordan and Williamson 1985).

[4]For a reflection on the legal and social constitution of the right-holder, with specific reference to contemporary science, see Kahn (1999).



If a contract is established between two firms for one firm to supply a key component of another firm's product, this contract in time binds the two firms together in a form that would prove costly to break even if the obligations of the contract themselves were suddenly removed. Here the rules of the game - the socially instituted contract between two firms - transforms the differential costs associated with different categories of transaction even though there is no necessary prior material basis for the institution of the contract. Transaction costs determine the rules of the game, *and* vice versa.

A Williamsonian approach to the analysis of institutions and institutional change can therefore usefully be understood as dividing transaction costs into two categories - those transaction costs that are *exogenous* to the rules of the economic game under analysis, and those that are *endogenous* to those rules. Shifts in transaction costs associated with, for example, technological change can typically be understood as exogenous to the analysis; transaction costs that are themselves determined by the actions of the economic agents under analysis - in particular by the formation of contracts and other shifts in the 'rules of the economic game' - can be understood as endogenous to the analysis. Understanding the relationship between transaction costs and the rules of the economic game in this way accommodates *both* Coasean insights - the determination of the rules of the economic game by transaction costs, and the role of the rules of the game in instituting transactions themselves.

This thesis adopts this broad analytic approach. However, where 'transaction cost economics' often *begins* with transaction costs, 'privileging' exogenous transaction costs within its explanatory approach, this thesis adopts the opposite attitude. In this thesis I begin from the analysis of the 'rules of the economic game', and only then turn to the analysis of the role of transaction costs within those rules. In this, the thesis is compatible with the broad tradition of the new economics of science, which foregrounds the distinctive

'Mertonian' rules of the scientific game in its analysis of science as an institution. To this end, in the next section I begin to draw out the connections between the new economics of science and one of the most important 'rules of the game' approaches within new institutional economics - that of Elinor Ostrom.

3.3.1 Elinor Ostrom's Institutional Analysis and Development approach

Ostrom's contributions to new institutional economics are wide-ranging: substantively, Ostrom has pioneered the study of polycentric common pool resource governance institutions, drawing economists' attention to a category of economic governance institution that had hitherto often been neglected; methodologically, Ostrom's 'Institutional Analysis and Development' framework has provided an analytic framework for economists engaged in the qualitative study of economic institutions; metatheoretically, Ostrom has developed a set of high level resources for theorising the 'rules of the economic game'

The core of Ostrom's project is her work on commons governance institutions - institutions which manage common-pool resources like fisheries, irrigation systems, pastures and forests, where over-exploitation of the resource is a danger, and access to the resource is difficult or costly to restrict. In her influential book *Governing the commons* (1990), Ostrom discusses a wide range of case studies of local 'polycentric' commons governance systems - cases where local communities have self-organised governance structures involving neither top-down state control nor private property rights: communal land management in small communities in Switzerland and Japan; irrigation systems in Spain and the Philippines; groundwater basin management in the US; and others.

In analysing such case studies, the primary focus of Ostrom's work is empirical - her goals are, first, to demonstrate through empirical example that many common frameworks for understanding commons governance are flawed or limited (e.g. Hardin 1968; Olson 1965), and second, to understand which institution-design principles are, in practice, most effective in managing common pool resources.

Ostrom's Institutional Analysis and Development (IAD) framework was initially developed to assist researchers studying common pool resource gov-

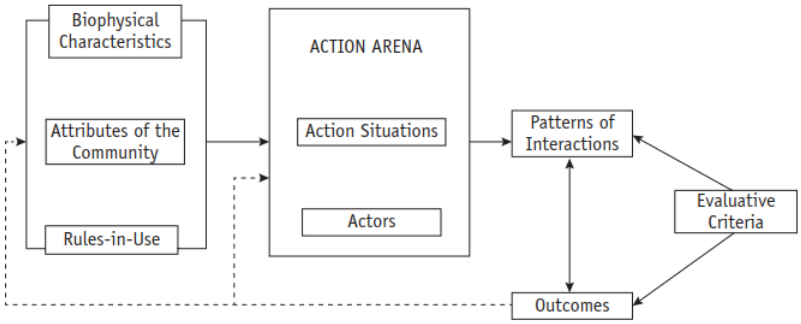


Figure 3.1: The Institutional Analysis and Development framework, as depicted in Strandburg et al. (2017), p. 15, after Ostrom (2005). Licensed under a Creative Commons Open Access licence CC-BY-NC-ND 4.0. To view a copy of this license, visit https://creativecommons.org/licenses/by-nc-nd/4.0.

ernance institutions of this type. Figure 3.1 schematises the key features of any institution in the IAD approach's terms.

Researchers operating within this framework collect data on the biophysical characteristics of the common pool resource in question; the attributes of the community, including the boundaries of its membership and the roles occupied by members; and the rules that govern the community, including the rules for transforming those rules. These features of the community, taken together, determine the 'action arena': the natural and social constraints that dictate the range of possible actions, and the incentives associated with those actions, for members of the community. Actors then make decisions within this action arena, which generate outcomes that may in turn transform the institution's structure. Observers such as researchers can analyse these actions and outcomes, to evaluate the success of the institution as a whole using their chosen evaluative criteria.

Within these categories of observation, Ostrom and her colleagues detail many more specific elements of institutions for researchers to observe. For example, in specifying the rules-in-use associated with an institution, Ostrom suggests that researchers should specify what sanctions are available to disincentivise transgression of the institution's rules; which members of the community have the authority to wield those sanctions; under what circumstances the sanctions are applied, etc. Similarly, when analysing the

attributes of the community, researchers should be able to specify what criteria must be met for an actor to be a member of the community; what entitlements and obligations are associated with different forms of community membership; under what circumstances community members may be expelled from the community, or have their role within the community transformed; etc.

Ostrom's IAD approach cannot immediately be applied, in its canonical form, to scientific institutions: the difference between a common pool natural resource, and the knowledge resources used and created by scientific researchers has implications for how the IAD approach can be applied to science. In some of her later work, Ostrom began to extend the IAD framework to knowledge commons (see in particular Hess and Ostrom 2007), and this project has since been continued by others. Most notably, Frischmann, Madison, Strandburg and others (Frischmann, Madison, and Strandburg 2014; Madison, Frischmann, and Strandburg 2009a,b) have developed a series of modifications to the IAD approach which, they argue, enable researchers studying scientific institutions to apply an amended IAD framework to the analysis of scientific commons. I will again return to this framework in more depth Part II.

For the purposes of this chapter, what matters is that Ostrom's Institutional Analysis and Development approach provides a methodological framework for the qualitative analysis of the 'rules of the scientific game'. These rules of the game in turn determine the *endogenous* transaction costs associated with a broader transaction costs analysis. This approach will become particularly important in Part III of the thesis, where I will discuss the impact of developments in information technology on the transaction cost economics of Mertonian and 'open' science.

To summarise the argument of this chapter so far: new institutional economics can be schematised into 'old' and 'new' institutional economics; new institutional economics can in turn be schematised into approaches that emphasise transaction costs, and those that emphasise the 'rules of the game'. Transaction costs and rules of the game are in many cases, and in practice, 'co-constituting'. Nevertheless, this thesis primarily falls within the latter subtradition: my emphasis in later chapters will be on the rules of the scientific game, and how they shape scientists' day-to-day practice. In order to describe these 'rules of the scientific game', I will be making use of Elinor Ostrom's Institutional Analysis and Development framework, as modified by Madison, Frischmann, Strandburg, and others for the application to knowl-

edge institutions. However, this framework only provides the resources to understand the *qualitative* structure of scientific institutions. The thesis will also develop a set of *formal* models of scientific institutions, and in the next section I look in more depth at some of the resources I will be drawing on to do so.

3.4 Institutions as games and networks

In the previous section, I discussed institutional economics, with a particular focus on Elinor Ostrom's Institutional Analysis and Development framework. Institutional economics in general seeks to embed traditional economic analysis within a broader, historicist perspective, which emphasises that the incentive structures within which economic actors operate, and the behavioural dispositions that guide them, are historically variable. Ostrom's IAD framework provides one systematic way to specify the historically variable 'action arenas' within which economic agents' actions take place

In chapter four I will draw on Ostrom's IAD approach, to discuss the social construction of the scientific action arena. As I discussed in chapter two, however, existing work in the economics of science has already proposed detailed accounts of the incentive structure generated by scientific institutions: the 'rules of the game' of scientific practice.

In particular, the 'new economics of science' suggests that scientific production can usefully be modelled using the tools of game theory as an n-player prisoner's dilemma, while one strand within the 'economics of scientific knowledge' suggests that veritistic social epistemology provides a useful approach to understanding the *epistemic* outcomes of scientific practice. Agent-based modelling provides a formal toolkit within which the 'evolutionary' dimensions of both of these dimensions of scientific practice can be explored.

In this section, I will look in more detail at some of the modelling approaches associated with these analytic frameworks. In section 3.3.1, I look at game theoretic analyses of the prisoner's dilemma, with a particular emphasis on evolutionary game theory. Then in section 3.3.2, I look at opinion dynamics modelling, with a particular emphasis on variants of the voter model. These two categories of model provide two different ways of representing the 'action arena' within which scientific research takes place: one focussed on the question of cooperation between scientific actors, and the other on the

question of the credibility of scientific output.

3.4.1 Science as non-cooperative game

In Chapter Two I discussed Dasgupta and David's proposal to model Mertonian science as a prisoner's dilemma. The basic two-player prisoner's dilemma is a two-player noncooperative game of the form:

	C	D
C	x, x	z, w
D	w, z	y, y

Where $w > x > y > z$

Or, presented differently:

	C	D
C	$a - c, a - c$	$-c, a$
D	$a, -c$	$0, 0$

Where $a > c > 0$

As I discussed in chapter two, in the basic one-round version of this game the only Nash equilibrium is *defect*, *defect*. Despite the fact that both players gain higher payoffs from universal cooperation than from universal defection, cooperation is a strictly dominated strategy. I likewise discussed, in chapter 2, some of the ways in which researchers have 'solved' the prisoner's dilemma, within classical game theory, via the introduction of iterated play. In this section I will briefly discuss some of the alternative analytic frameworks that researchers have used to explore the problem of the dilemma. In particular, two subtraditions have departed from the traditional game-theoretic conception of rational payoff-maximising strategic behaviour: *behavioural* and *evolutionary* game theory.

On the one hand, *behavioural* game theory has emphasised some of the ways in which the behaviour of real economic actors does not match the traditional assumptions of self-interested payoff-maximisation rationality. In theoretical work, researchers have explored a range of alternative decision and payoff functions: heavier weighting of losses relative to gains than predicted by traditional expected utility theory (Tversky and Kahneman 1992); the replacement of purely self-interested payoff functions with other-oriented

payoffs (Loewenstein, Thompson, and Bazerman 1989); preferences for fairness in game outcomes (Rabin 1993); and others. At the same time, researchers have explored how real world economic actors make strategic decisions in game-theoretic scenarios, carrying out empirical research both in ethnographic field-site studies (e.g. Henrich 2004), and in laboratory experiments (e.g. Chaudhuri 2011). Both these strands of empirical work have opened game theory to a range of possible behavioural dispositions beyond the traditionally 'selfish' payoff maximising agent.

At the same time, some strands of *evolutionary* game theory have explored the role of environmental (social) selection mechanisms, rather than strategic decision-making, in game dynamics. The term 'evolutionary game theory' can be used to denote two different approaches to game-theoretic modelling. On the one hand, evolutionary game theory can denote the use of classical game theory to model evolutionary outcomes, an approach pioneered by John Maynard Smith and George R. Price (Smith and Price 1973; Smith 1988). The difference between this form of evolutionary game theory and traditional game theory is primarily a difference of interpretation rather than method. Thus in evolutionary game theory in this sense, a mixed-strategy equilibrium can be interpreted as an 'evolutionarily stable strategy': rather than a probabilistic weighting of pure strategic choices, a mixed-strategy equilibrium is here understood as the ratio of pure strategies distributed across a large population.

At the same time, 'evolutionary game theory' can also denote an approach that emphasises the dynamics via which strategies are selected among a large number of players in a population. In this approach, players with more successful strategies in a given game are selected for, and those with less successful strategies are selected against. This can be understood as the elimination of less successful agents and their replacement by more successful ones; as the decision by persistent agents to replace less successful strategies with more successful strategies, via a social learning process; or both. In any of these cases, models of this kind depict a dynamic process of strategy selection, in which selection across variation leads to the spread of some strategies and the reduction in incidence of others.

The pioneering example of this approach is Axelrod (2006). Axelrod, taking as his illustrative game the traditional two-player prisoner's dilemma, ran a tournament in which different game strategies competed against each other in a series of simulations. Axelrod found that the strategy 'tit for tat', in which players begin the game playing cooperatively, but thereafter adopt

their opponent's previous strategy, to be the most evolutionarily successful strategy. The major contribution of Axelrod's book was not this finding, however, but his method. The idea of running simulations to track the dynamics of strategy selection has become a major strand of evolutionary game theory, and is the subtradition within which the game-theoretic models used in this thesis should be located (Perc et al. 2013).

Clearly, this kind of evolutionary approach is not appropriate to many strategic scenarios. In the kind of 'one-off repeated interactions' that characterise many strategic scenarios analysed by game theorists, there is no broad population of actors within which strategies can be selected. In analysing situations of monopolistic competition, for example, game theory is valuable precisely because of the very small number of strategic actors, and the use of evolutionary game theory is inappropriate.

Nevertheless, there are many scenarios in which selection of strategies within very large populations seems like an appropriate framework for the analysis of strategic behaviour. One of the assumptions of this thesis is that the analysis of large-scale scientific research communities is one such scenario. Contemporary scientists are part of a very large - global - community of practice. Moreover, scientists' reputational success or failure is in large part 'selected for' by the collective judgements of that community. Understanding science in this way furthermore connects to the broader qualitative and philosophical tradition of evolutionary analysis of science, which I discussed in chapter two. In chapter five, I therefore adopt an evolutionary game-theoretic approach to the analysis of scientific 'strategies'.

The above discussion of different game-theoretic approaches to the analysis of strategic interactions takes the two-player prisoner's dilemma as its starting point. As discussed in chapter two, however, for Dasgupta and David, and for the new economics of science more broadly, most scientific research is best represented as an n-player, rather than a two-player, prisoner's dilemma.

I postpone a fuller discussion of the n-player prisoner's dilemma to chapter five. Here, I will only note that the same theoretical decisions that must be made in relation to the analysis of the two-player prisoner's dilemma, also apply to the analysis of the n-player dilemma - questions I will return to in Part II.

3.4.2 Science as epistemic network

In the previous subsection I discussed how game theory - specifically, games within the family of 'prisoner's dilemmas' - can provide a toolkit for modelling scientific action situations around cooperation or noncooperation. This approach is well suited to the strategic decisions scientists must make around input sharing. However, the choice between cooperative or noncooperative behaviour is not the only dimension of scientific practice relevant to the problems discussed in this thesis. This thesis is concerned not only with the problem of *cooperation*, but also with the problem of *credibility* - the question of the epistemic outcomes associated with different scientific institutional structures.

In chapter two I discussed a range of different approaches taken to this question within science studies, and focussed on one in particular: veritistic social epistemology. I further discussed the use of agent-based modelling to simulate the evolutionary dimensions of scientific research dynamics. In this sub-section I will explore one particular modelling approach that can potentially illuminate these issues: opinion dynamics modelling.

From veritistic social epistemology to opinion dynamics modelling

In section 2.5.2 I discussed veritistic social epistemology as a framework for both analysing and evaluating the epistemic consequences of different institutional structures. Veritistic social epistemology assumes a broadly realist epistemological framework - though, as I discussed in more detail in section 3.6, this assumption need not require commitment to any strong form of metaphysical realism. Moreover, veritistic social epistemology assumes that epistemic delegation is central to opinion formation: any given social actor heavily relies on the judgements of other social actors, such that individuals' opinions are socially, not simply individually, determined. Finally, veritistic social epistemology assumes that the toolkit of economic modelling can provide valuable resources for comparing different social structures and practices, and their epistemic outcomes.

Within this basic framework, however, a range of different modelling approaches are available. In section 2.6 I discussed agent-based approaches to the analysis of science as an epistemic system. In this section I will discuss the growing sub-field of opinion dynamics modelling, the category of epistemic

modelling adopted in this thesis.[5] When Goldman proposed veritistic social epistemology as a theoretical framework, opinion dynamics modelling was in its infancy. With the growth of interest in social networks (Benkler 2006; Jackson 2010), and the increasing use of computer simulation in social science, however, opinion dynamics modelling has become a major subtradition within social epistemology. As I will argue below, this modelling approach - already adopted in early form by David (1998) - is particularly well suited to the analysis of scientific institutions.

In the remainder of this subsection I provide an overview of the different available approaches to opinion dynamics modelling. This lays the groundwork for the discussion of specific opinion dynamics models, in chapters six and eight.

Approaches to opinion dynamics modelling

Opinion dynamics modelling is an approach to the analysis of epistemic communities which sees those communities as networks of agents, each of whom update their opinions in response to the opinions of neighbouring agents. Opinion dynamics modelling therefore draws on graph theory, for its analysis of network behaviour. But opinion dynamics modelling also makes extensive use of computer simulations to track the behaviour of epistemic systems. In this thesis, as with the use of evolutionary game theory, I will be using computer simulations to model scientific opinion dynamics.

Opinion dynamics modelling is grounded in the idea that individual agents exist in a social network, and modify their opinions via interactions with other agents in that network. In opinion dynamics modelling we thus specify the nature of the opinions, the network, and the agents' learning mechanisms, and observe - given these assumptions - the dynamics of the network as a whole. Opinion dynamics modelling is therefore part of a broader simulations-based literature, which studies how potentially complex macro-behaviours can be produced by relatively simple micro-behaviours (Tesfatsion and Judd 2006; Acemoglu and Ozdaglar 2011).

In modelling the opinion dynamics of scientific communities, we are interested in whether and how the scientific community as a whole reaches a collective opinion. As I will discuss in greater detail in Part II, one of the characteristic features of science as an epistemic system is that the opinions

[5]For recent overviews of opinion dynamics modelling see Acemoglu and Ozdaglar (2011) and Sîrbu et al. (2017).

of any one member of the community, however authoritative, do not bear this authority in themselves. On the contrary, the epistemic authority we grant to scientific claims is based on the fact that these are in some sense *collectively* determined opinions. Scientific epistemic authority is grounded in an ongoing discursive process in which individual scientists' knowledge claims are subject to rational and evidence-based evaluation and contestation by others. Only scientific claims that are broadly endorsed, as an outcome of this process, merit the full social authority of science. It is therefore important for those studying scientific opinion formation to study the opinion dynamics of the scientific community as a whole, not just, for example, the scientific method that informs individual scientists' research practice. Opinion dynamics modelling provides us with a set of formal resources for modelling this collective process.

The components of an epistemic system

As just outlined, opinion dynamics modelling studies epistemic systems as *networks* of *agents* updating their *opinions* via a specific *learning process*. In order to describe a specific opinion dynamics model, therefore, we need to specify how we are modelling each of these components of the system. A wide range of different opinion dynamics models, with very different properties, are available, depending on these decisions about model components. At present there is little consensus in the literature around which modelling assumptions are most appropriate to which real-world empirical cases. In chapters six and eight I will propose specific models to capture elements of specific scientific institutions. In the remainder of this subsection, I will outline the 'space of possibilities' presented by the range of options available for each of these modelling elements.

A *belief* held by an agent can be modelled as a variable which is updated by a learning process. Such belief variables may be discrete or continuous. Discrete opinions are most commonly represented as a binary choice, such as the choice between belief or non-belief (e.g. Mobilia, Petersen, and Redner 2007). Alternatively, beliefs may be modelled as a continuous range of possible values. This range of values may represent a range of possible beliefs - for example, an agent's political alignment on an ideological spectrum (e.g. Ben-Naim 2005) - or it may represent the strength of conviction or probabilistic weighting assigned to a single belief - as in Bayesian probabilistic weighting of opinions (e.g. Martins 2009). These approaches can also be combined, as

in the CODA model described in chapter seven, by treating the binary choice to accept or reject a proposition as a decision made on the basis of a continuous probabilistic estimate of the proposition's truth or falsehood (Martins 2008). Finally, although this option remains underexplored in the literature to date, agents may be modelled as holding multiple beliefs. In this scenario, agents' update rules may consider not only neighbours' beliefs, but also the relationship between agents' own beliefs.

The next important component of any opinion dynamics model is the *network* that links agents. A wide range of network structures are, naturally, available. Agents may exist in a basic unstructured graph, with each agent linked to a given number of neighbours. Agents may exist in a structured graph, with some agents linked to more neighbours than others, or with weighted or directional links between agents. Thus, for example, some recent models simulate the role of the mass media by establishing a small number of agents with unidirectional links to all or most other agents in the network (e.g. Quattrociocchi, Caldarelli, and Scala 2014). These agents are highly influential, relative to most agents with only a small number of links to neighbours. Some recent researchers have created simulations with dynamic graphs, in which links between agents are created, destroyed and weighted based on interactions between agents (e.g. Proskurnikov, Matveev, and Cao 2016).

Having established agents with beliefs within a network, the next component of any opinion dynamics model are the *update rules* via which opinions are transformed. What update rules are available depends on other features of the model. In the case of binary beliefs, for example, there are a limited range of ways in which an agent may update their beliefs in any given round of play. In a so-called 'voter model' (Clifford and Sudbury 1973; Holley and Liggett 1975) an agent will adopt the belief of one of their neighbours, selected at random. In 'majority rule' models agents will survey their neighbours and adopt the belief held by the majority of agents in their neighbourhood (e.g. Galam 2002). Or we may model some agents as 'contrarians', who adopt the *minority* belief in their neighbourhood (e.g. Crokidakis, Blanco, and Anteneodo 2014).

In systems where agents hold continuous opinions, a range of update rules are likewise available. Agents may adopt a Bayesian update rule, either adhering to 'full Bayesianism', which requires probability estimates for all elements of the system, or a 'rule of thumb' Bayesianism, which simply fiats an estimated likelihood that neighbouring agents' opinions are correct

(Acemoglu and Ozdaglar 2011). Alternatively, agents may adopt a range of alternative 'rule of thumb' update rules. Agents may, for example, update their beliefs to the mean of their neighbourhood's beliefs (DeGroot 1974). Agents may update their beliefs conditionally - for example, only attending to the beliefs of neighbours if those beliefs are close enough to their own (Hegselmann, Krause, and others 2002). If agents exist on a structured - or dynamic - graph then the weighting of links to neighbours may naturally influence the weighting accorded to neighbours' beliefs in agents' update rules. Moreover, different agents within the same model may have different update rules. A model may, for example, include a subset of 'fanatics' who never update their opinions regardless of their neighbours' beliefs (Stauffer and Sahimi 2007).

Finally, it is possible to establish opinion dynamics models in which agents' update rules are also influenced by elements outside the network. In some sense, this is a definitional rather than a substantive point, since an 'external' influence can always be modelled as 'internal' through the creation of an appropriately weighted and directional graph. Nevertheless, it is common for some epistemic systems to treat agents' observation of 'nature' as an exogenous influence from outside the epistemic network, often using the paradigm of the Ising model in physics (Galam, Gefen, and Shapir 1982). In this thesis I will be considering scientific agents who observe external nature. This nature therefore represents a potential external influence on agents' belief update rules.

This survey provides a brief overview of some of the possibilities of opinion dynamics modelling. Clearly, a wide range of possible approaches are available, with little consensus as to the most appropriate modelling approach. In this thesis, I will therefore primarily take my guidance from previous scholars. My goal will be to build on major existing opinion dynamics models of scientific research within the new economics of science, in order to examine the consequences of open inputs science within those models. In this way the models in the thesis remain tethered to the existing sub-tradition, while building on and departing from it in a set of specific ways.

3.4.3 Summary

In section 3.3 I discussed Ostrom's approach to the qualitative analysis of an 'action arena' - a social environment within which agents make decisions constrained by local social norms and sanctions. In this section, I have

explored two theoretical toolkits for the formal modelling of agents' behaviour *within* such an action arena. First, I looked at game theory as a toolkit for analysing the likelihood of cooperative and non-cooperative behaviour. This way of modelling an action arena is strongly consistent with Ostrom's own work. In section 3.4.2 I then looked at *opinion dynamics modelling* - an approach for understanding the epistemic behaviour of a social system, given that system's structure and its members' learning behaviour. As I will explore in later chapters, such an opinion dynamics system can itself be understood as an action arena: an epistemic behavioural space created by contingent historical institutional norms and practices.

In summary, then, the two dimensions of scientific practice with which this thesis is concerned - cooperation and credibility - can be represented using two different modelling toolkits, bound together by the qualitative analysis of the action arena that generates them. This is the approach I will be adopting in the chapters that follow.

In the remainder of this chapter, I will address some of the methodological and metatheoretical challenges associated with this approach.

3.5 Four concepts of modelling

One of the core methodological challenges confronting practitioners of formal economic modelling is the question of the status and explanatory function of economic models. As is almost universally acknowledged, formal economic models are in many ways highly unrealistic: the assumptions embedded in most economic models depart in many ways, both large and small, from the complex reality of the social phenomena they aspire to represent.

For many critics of economics, this lack of realism in economic modelling is a serious, for some fatal, challenge to the scientific status of economics as a whole, or of some important subset of economic theories. The sociologists Hirsch, Michaels and Friedman, for example, claim that

> economists leave themselves no mechanism for learning about the crude and messy empirical world that so defies their models. Economists pay a heavy price for the very simplicity and elegance of their models: empirical ignorance, misunderstanding, and, relatedly, unrealistic and bizarre policy recommendations. (Hirsch, Michaels, and Friedman 1987: 320)

Such sweeping critiques are regarded with some scepticism by most economists. Nevertheless, even within the discipline of economics, some scholars have expressed concerns about the realism of many formal models (see for example Rubinstein 2006). What is the methodological warrant for the use of apparently highly schematic and unrealistic formal models to analyse a complex economic and social reality?

Different economists and philosophers of economics have answered this question in different ways. In this section I will categorise these methodological or metatheoretical strategies into four groups: accounts of models as predictions, deductions, metaphors and caricatures. I will argue that the models in this thesis are best understood in terms of the last of these categories of methodological rationale: models as caricature.

One influential defense of unrealistic models is that proposed by Milton Friedman in his 1953 'The methodology of positive economics'. In that paper, Friedman argues that the only scientifically relevant property of economic models is their predictive power. In Friedman's words

> The ultimate goal of a positive science is the development of a 'theory' or 'hypothesis' that yields valid and meaningful (i.e., not truistic) predictions about phenomena not yet observed. (Friedman 1953: 7)

Even more strongly:

> The only relevant test of the validity of a hypothesis is comparison of its predictions with experience. (Friedman 1953: 9)

This can be seen as the predictive warrant for methodological lack of realism.

An alternative approach has been to justify unrealistic models on *deductive*, rather than predictive, grounds. One important early advocate of this approach is the 19th century political economist and philosopher John Stuart Mill (Mill 1877), who argues that political economy is an *a priori* science: the basic assumptions of the science of political economy are legitimated, for Mill, by their foundation in our general experience of one dimension of human nature, rather than by their empirical validation. This broad category of argument has been reformulated by more recent theorists; for example,

Hausman (1992) argues that a version of Mill's argument is the tacit methodological approach behind the centrality of neoclassical equilibrium analysis in much post-war economic theory.

A third approach is to emphasise the *rhetorical* dimension of economic models. For McCloskey (1998), the most influential advocate of this methodological strategy, economic models should be understood as forms of rhetoric, in two senses. First, economic argument should not be understood in terms of proof - whether verificationist, falsificationist or deductive - but rather in terms in persuasion; second, paradigmatically scientific discursive tools such as hypothesis testing and mathematical modelling should themselves be understood as literary techniques. In particular, McCloskey argues that economic models are "nonornamental metaphors" (McCloskey 1998: 40). This broad approach to understanding economic models as continuous with literary technique has been adopted by a range of other economic theorists, with different scholars variously describing models as "fables" and "tales" (Rubinstein 2006, 2012); "stories" (Morgan 2001); "parables" (Cartwright 2010); "analogies" (Gilboa et al. 2014), "allegories" (Highfill 1992), and other similar terms.

The final approach to economic methodology understands economic models as, in Gibbard and Varian's words, 'caricatures' (Gibbard and Varian 1978). For this approach, the purpose of models is to formally capture some core elements of social and economic behaviour, while disregarding others. These 'caricatural' models, however, do not aspire to capture just one, central, invariant element of human behaviour. Rather, caricatural models aspire to capture a specific dimension of a given economic situation, where the elements of the economic situation under analysis are contextually determined by the modeller. Unlike Mill's deductive approach, therefore, the 'caricatural' approach is associated with a *pluralist* attitude to modelling. From this pluralist perspective, different modelling approaches may be appropriate for analysing different economic situations. Moreover, different modelling approaches may be appropriate for analysing the *same* economic situation, if a different element of that situation is to be analysed.

For this approach, in Gibbard and Varian's words:

> Often the assumptions of a model are chosen not to approximate reality, but to exaggerate or isolate some feature of reality. (Gibbard and Varian 1978: 673)

This 'caricatural' approach to economic explanation is the approach adopted

in this thesis. The models in this thesis do not aspire to strong predictive power, as Friedman's methodology requires; neither do they aim to deduce their findings from *a priori* features of human beings as economic creatures, as does Mill's. The models proposed in this thesis are rhetorical in nature - as McCloskey suggests - but this rhetorical character is part of a broader rational, empirically-oriented social scientific discursive practice. Within such a discursive practice, the 'caricatures' of formal models can perform important *explanatory* functions. As Gibbard and Varian put it:

> The theorist's approach to his problem is not to try for the closest tractable approximation of this complex reality, but to tell a simple story that captures some of its features.... The model poses the question, what would happen if this story were true? (Gibbard and Varian 1978: 674)

In this thesis, then, I suggest that the methodological warrant for the formal models developed in later chapters should be understood in terms of the following set of independent, but connected, ideas:

- many economic models primarily serve an explanatory function;
- this explanatory function operates by isolating and schematically representing specific elements of social reality, while disregarding others;
- models thereby illuminate one set of causal mechanisms at work in economic life, while neglecting other mechanisms;
- the accuracy of the model depends on the relative importance, in the specific phenomena to which it is applied, of the causal mechanisms it highlights;
- different models can legitimately focus on different causal mechanisms within the same social phenomena;
- the same social phenomena can therefore legitimately be represented by different models;
- analysts must therefore make a pragmatic, qualitative, and context-dependent decision about which modelling approach is best suited to the analysis of any given economic phenomenon, for any given analytic purpose;

- this pragmatic and qualitative decision must be informed in part by factors that fall outside the scope of the models themselves.

The models proposed in this thesis, I suggest, should be evaluated in these terms. They provide *one* window onto the analysis of scientific practice. Other analytic approaches would provide other windows, and other insights. The claim of this thesis, however, is that the insights into the structure and dynamics of scientific institutions proposed in the chapters that follow can make a contribution to the broader, collective project of the 'new economics of science'.

3.6 Epistemology and science studies

One final metatheoretical issue should be addressed, before I turn to the substantive argument of Part II: the epistemological commitments involved in the social-scientific analysis of scientific production. As I discussed in chapter two, in the latter part of the twentieth century, a series of debates over the 'social construction' of science played out across a range of disciplines.[6] These debates in part focused on social-scientific methodology and analytic strategy, but were also often motivated by epistemological or philosophical concerns around relativism, realism, and the legitimacy of scientific, including social-scientific, knowledge.

Very broadly speaking, these debates can be understood in terms of the distinction between constructivist and anti-constructivist understandings of the *content* of scientific knowledge. While practitioners of the social science of science are broadly agreed that science has many important social dimensions, there are substantial disagreements about the explanatory reach of social science with respect to the substance of scientific claims themselves. More 'constructivist' social scientists (e.g. Bloor 1976) argue that the content of scientific knowledge can be explained in social terms; more 'realist' social scientists (e.g. Popper 1959) argue that, although social practice has important influences on the content of scientific knowledge, non-social dimensions of epistemology are essential to understanding the epistemic warrant of

[6] On 'social constructionism' in the social sciences in general, see Berger and Luckman (1991), Searle (1995) and Hacking (1999); on social constructionism in economics specifically, see Mäki (2002).

scientific truth-claims.[7]

The epistemological or philosophical dimensions of these debates are complex, and I will not review them here. In Law and Pepperell (2017), I discuss these debates in more detail, and propose a specific metatheoretical position. There I argue that the recent work of the pragmatist philosopher Robert Brandom provides a route through this set of issues that allows us, as social scientists, to preserve the strengths of both constructivist and realist approaches. Just as the work of David Lewis on the game-theoretic analysis of social convention (Lewis 2008) has been influential in economics, not just within philosophy, so, I argue, the work of Robert Brandom, which extends Lewis's analysis of social convention into a complex linguistically-focused 'normative pragmatics', can potentially also provide resources for a formal analysis of normative, institutional structures within the social sciences more broadly (Brandom 1994, 2002, 2008, 2009a,b).

For the analysis carried out in this thesis, however, no strong metatheoretical commitments are required. In terms of its social epistemology, and as already discussed in chapter two, this thesis presupposes two things: first, that the content of scientific research is indeed determined, in part, by social practices and institutions. Second, that the output of scientific research can nevertheless also be normatively evaluated in terms of its 'fit' to the objective state of affairs in nature, however we choose to understand the latter category.

As I discussed in chapter 2, this latter assumption requires at least 'weak realism', of some form. Nevertheless, the thesis remains broadly agnostic between a range of epistemological or metatheoretical commitments. What matters, for the analysis carried out in this thesis, is a minimal set of commitments to normative epistemology: scientific knowledge is shaped by the norms and practices of scientific institutions, and those scientific institutions can in turn, in principle, be evaluated normatively in terms of the epistemic value of the knowledge they produce.

[7]For a comprehensive survey of post-Kuhnian sociology of science, see Zammito (2004). For a more recent overview of the economics of scientific knowledge, see Zamora Bonilla (2012).

Part II

Chapter 4

The rules of the scientific game

4.1 Introduction

In part I of the thesis I did several things. First, I outlined the task of the thesis: the comparative institutional analysis of closed and open inputs science. I discussed the two dimensions along which the thesis will compare these institutional forms: cooperation in the scientific production process, and reliability of scientific outputs. I located these issues within two key theoretical traditions: the new economics of science and the economics of scientific knowledge. Finally, I outlined the tools I will use to carry out that task: a 'rules of the game' new institutionalist approach to the description of scientific institutions, and agent-based modelling of the behaviour of scientists given those rules.

The remainder of the thesis will carry out this project. My analytic strategy, across the chapters that follow, has several stages:

First, in this chapter, I will outline a simple qualitative analysis of the 'rules of the scientific game'. The goal here will be to build on the work of Dasgupta and David, to give an account the rule-bound structure of science as an institution in a manner sufficiently general to permit the development of formal models of scientific research dynamics, in later chapters.

I will then use the qualitative discussion of this chapter to inform the construction of several distinct agent-based models of scientists' actions, in chapters five and six. In those chapters I will develop two sets of models, focused on the two dimensions of scientific institutions that I discussed in part I. In chapter five, I will develop an agent-based evolutionary game-theoretic

model of cooperation in intermediate input sharing within the traditional 'Mertonian' scientific institutional structure. Then, in chapter six, I will develop a set of opinion dynamics models of the impact of scientists' behaviours on the epistemic outcomes of science as a system.

These models taken together provide a composite picture of the structure and dynamics of traditional 'Mertonian' science. In Part III of the thesis I will then use these models as the starting point for a *comparative* analysis of 'Mertonian' and 'open inputs' science.

The core task of Part II of the thesis, then, is the development of a set of interlinked models of traditional, Mertonian science. I begin this task in this chapter by embedding Dasgupta and David's account of Mertonian science within a more expansive qualitative account of science as an institution.

4.2 Science as a global institution

In Part I of the thesis I discussed science as an institution in Mertonian terms. For Merton - and, building on Merton, for Dasgupta and David's new economics of science - science as an institution has two key features: a set of community norms - universalism, communism, disinterestedness, and organised scepticism - and a set of incentive structures based on reputational rewards for priority of publication. Dasgupta and David formalise this latter element of science's institutional structure in their model of science as n-player prisoner's dilemma.

Although this basic Dasgupta and David model serves as the core of my analytic approach, there are important elements of science as an institution that this simple n-player prisoner's dilemma model does not capture. What determines the size and membership of the community of players who together compose the n of any given n-player game? How does any given game relate to the institution of science as a whole? What is the internal structure of the scientific community within which such games are played? These are questions that are not answered by the simple Dasgupta and David model, and which this chapter begins to address.

To tackle these issues I draw on two key sets of additional resources. First, in section 4.2.1, I draw on Michael Polanyi's account of the 'republic of science'. Polanyi's concept gives an account of the macrosocial structure of science as a system that will inform the remainder of the thesis: science as a polycentric governance institution, the epistemic output of which is best

understood as a set of *emergent effects* rather than as aggregated products of distinct knowledge production processes. Then, in section 4.3.1, I draw on Jason Potts' concept of the *innovation commons*, as a novel way to understand important subcomponents of this macrosocial system. Taken together, these theoretical concepts give us a more complex picture of the structure of science as a global institution.

4.2.1 The republic of science

Michael Polanyi's contributions to the science of science can broadly be schematised into two categories. *First*, Polanyi is notable for an emphasis on the importance of *embodied knowledge* - 'personal knowledge' or the 'tacit dimension' of knowledge, as he put it in two of his major works (Polanyi 1958, 1967). Polanyi, like a range of twentieth century theorists (Wittgenstein 1958; Garfinkle 1967; Pickering 1992), was concerned to rebut the common understanding of knowledge as something necessarily detachable from any specific instantiation. For Polanyi, the way we know is, in the first place, personal and physical - we know *how* before we know *what* (cf. Brandom 1994).

Second, Polanyi is concerned with the way in which science is organised as a social system. Here Polanyi is concerned not with the personal dimensions of science, but with the elements of science that are characteristically and necessarily transpersonal, and that therefore exceed any individual's tacit or explicit intention. These dimensions of science's institutional structure can, for Polanyi, only be analysed and understood in *macrosocial* terms, by studying science as a *system*.

Polanyi's preoccupation with these macrosocial characteristics of science emerged from a series of debates among social theorists, scientists, and policymakers in the middle decades of the twentieth century, concerning the optimal mode of organisation of science as a system. For one influential strand of thought in this period, the major shortcoming of actually existing modes of scientific organisation was their relative lack of hierarchical and centrally planned social organisation. From this perspective, the social task of science is to serve society as a whole: the needs of society, and thus the preferred focus of scientific research, can best be identified by central planners whose decisions can direct the research funding and managerial structures of a broader scientific system. This perspective was most influentially articulated by J.D. Bernal in his *The social function of science* (1939), a work that

came to be seen as a foundational text in the emerging subfield of the social studies of science.

Bernal and Polanyi both agree, like Vannevar Bush (1945), that scientific research requires substantial government subsidy to enable the pursuit of research projects that would otherwise be underfunded by market mechanisms. Where these figures disagree is in their understanding of the optimal internal structure of science as an institution. Where Bernal sees science and society as best served by a relatively hierarchical internal structure of scientific institutions, Polanyi argues that a decentralised, polycentric internal institutional structure is essential to the epistemic pluralism required for positive scientific outcomes.

For Polanyi, in other words, and for a range of other scientists of the time associated with *The Society for Freedom in Science*, Bernal's vision of science was to be opposed on both political and scientific grounds (McGucken 1978). Politically, Polanyi and his colleagues were concerned that Bernal's vision of science tended towards a communist authoritarianism that would unacceptably restrict individual freedoms (Polanyi 1945). Scientifically, Polanyi was concerned that constraints on scientific freedom in the name of social planning would have negative *epistemic* impacts on scientific knowledge production. By subordinating truth-seeking and intellectual diversity to social planning, Polanyi argued, the fallibilist and pluralist dimensions of the scientific research process would be undermined. For Polanyi, as for Popper (1959), the internal diversity of scientific opinion was crucial to the evolutionary dynamic of selection among variation, which allows diverse scientific opinions to be appropriately evaluated against and alongside alternative perspectives.

From this evolutionary perspective a centrally planned epistemic economy, in which scientific research is directed by decision-makers in the interests of the perceived common good, is corrosive of scientific epistemic reliability and credibility. By contrast, a scientific epistemic economy driven by the local knowledge and preferences of many diverse scientific actors would in aggregate generate more reliable scientific knowledge - and thus also be of greater social value. Like a capitalist market operating with the dynamics of Adam Smith's 'invisible hand', a scientific community that did not aspire to follow the advice of central planners seeking the overall common good, would in practice be more reliably generative of robust public good outcomes.

Polanyi presented the most developed articulation of this understanding of science as a system in his classic 1962 article "The republic of science:

Its political and economic theory" (Polanyi 2000; cf. Polanyi 1941). There Polanyi characterises science as a liberal order composed of self-interested, cooperative individuals whose coordinated action results in a general socially beneficial result that no individual could have planned. In Polanyi's words:

> self-coordination of independent initiatives leads to a joint result which is unpremeditated by any of those who bring it about. Their coordination is guided as by 'an invisible hand' towards the joint discovery of a hidden system of things. (Polanyi 2000: 3)

Attempts to institute centralised planning of socially optimal scientific activity are therefore, in Polanyi's view, anathema to the scientific research process:

> any authority which would undertake to direct the work of the scientist centrally would bring the progress of science virtually to a standstill. (3)

Polanyi's 'invisible hand' metaphor aligns the process of scientific research with Adam Smith's analysis of market society (Smith 1976: 477). But Polanyi does not aim simply to analyse science by analogy with the market. Rather "the coordinating functions of the market are but a special case of coordination by mutual adjustment" (4). Polanyi's goal in his 1962 paper is to describe the social mechanisms via which this coordination by mutual adjustment functions in the case of science.

Polanyi's account of science's coordination mechanism has three parts.

First, Polanyi emphasises the role of authority in science. Like Kitcher (1993), Polanyi is critical of the idea that science is characterised by the rejection of authority - rather, science is a complex social system of authority-claims. Scientists only gain admission to the scientific community if they are willing to accept the authority of the community in their training and credentialing. Furthermore, scientists cannot function as members of the community without accepting the authority of those more expert than themselves in areas outside their own specialisation. However, this willingness to accept the authority of the scientific community is tempered by the obligation, as a working scientist, to challenge authority through innovation in one's own area of expertise - which is, in turn, the basis of one's own authority. In Polanyi's words:

> The authority of scientific standards is thus exercised for the very purpose of providing those guided by it with independent grounds for opposing it. (6)

Second, Polanyi analyses how this authority-system allows scientists to adjust their research in light of others' findings. The publication of research constantly modifies, even if in minor ways, the set of propositions that form the background framework for each scientist's own research. Each scientist adjusts their own presuppositions in light of the published findings of others - and this ongoing process of mutual adjustment allows the system as a whole to transform the propositions that are generally accepted through a series of usually incremental, sometimes revolutionary, changes. For Polanyi, therefore, publication of research findings is the form of information transmission that functions equivalently to the price system in market society. (4)

Third, Polanyi analyses the network structure that allows this system of mutual adjustment to function. For Polanyi, it is important that the social structure of science is organised as a large set of distinct but partially overlapping areas of expertise. In Polanyi's words:

> while scientists can admittedly exercise competent judgement only over a small part of science, they can usually judge an area adjoining their own special studies that is broad enough to include some fields on which other scientists have specialised. We thus have a considerable degree of overlapping between the areas over which a scientist can exercise a sound critical judgement. And, of course, each scientist who is a member of a group of overlapping competences will also be a member of other groups of the same kind, so that the whole of science will be covered by chains and networks of overlapping neighbourhoods. (7)

This network structure of the scientific community means that collective judgements can be formed as emergent properties of the network as a whole, without any individual scientist needing to be competent in more than a small subset of the areas of expertise required for the formation of those judgements:

> Scientific opinion is an opinion not held by any single human mind, but one which, split into thousands of fragments, is held

> by a multitude of individuals, each of whom endorses the other's opinion at second hand, by relying on the consensual chains which link him to all the others through a sequence of overlapping neighbourhoods. (7)

This vision of science as a network formed of partially overlapping local neighbourhoods is the core of Polanyi's concept of the 'republic of science'. For Polanyi, it is this network structure that allows scientific cooperation to transform local innovation and broader respect for authority into a system that is as a whole both dynamic and reliably truth-seeking.

Polanyi is interested in a particular form of cooperation in this network system: the cooperation associated with adjusting one's background beliefs in light of the findings published by others in the network. In this thesis, however, I am interested not only in this epistemic dimension of science as a system, but also in the role that cooperation in intermediate input sharing plays within the system as a whole. In the next section, I therefore begin to look in more detail at practices of cooperation within this scientific network.

4.3 Commons within the republic of science

The previous section outlined the basic account of the macrosocial structure of science as an institution that informs the thesis. This understanding of the 'republic of science' suggests that science as an institution is a complex polycentric structure composed of many smaller subcommunities. This understanding of the 'republic of science', however, does not illuminate the structure of the smaller-scale subcommunities out of which this macrosocial institution is composed. Analysing the basic structure of these subcomponent communities is the focus of this section.

As I discussed in chapter three, one of the key resources available for the study of smaller-scale scientific institutions is the research project derived from Elinor Ostrom's work on studying knowledge commons. In particular, the Workshop on Governing Knowledge Commons headed by Dedeurwaerdere, Frischmann, Madison, Schweik, and Strandburg has carried out a growing number of empirical case studies of scientific commons institutions.[1] These case studies include the Human Genome Project (Contreras 2014), the Rare Diseases Clinical Research Network (Strandburg, Frischmann, and Cui

[1] http://knowledge-commons.net/

2014), the citizen science astronomical project Galaxy Zoo (Madison 2014), medical data sharing by neuroscientists (Larson and Chon 2017), biobank data collections (Boggio 2017), the US Food and Drug Administration's Sentinel Initiative (Abbott 2017), and a range of other cases.

These empirical studies of the Workshop on Governing Knowledge Commons are encompassed by a methodological and theoretical framework derived from Elinor Ostrom's Institutional Analysis and Development (IAD) framework. I discussed the Ostromian IAD framework in chapter three, but in that chapter I postponed discussion of the extension of this framework to the study of scientific commons. In the remainder of this section I will discuss this extension and modification of the IAD framework, and its relevance to the study of scientific commons.[2]

As I discussed in chapter three, the basic IAD framework - illustrated in figure 4.1 - has several core elements. Researchers working within this framework study the attributes of the resource being managed by a commons governance community, the *rules of the game* that structure an 'action arena' within which economic action takes place, the behaviour of economic actors within the constraints of this action arena, and the outcomes of this behaviour, evaluated against the researcher's criteria for effective resource management.

Of these elements of the IAD approach, the most crucial is the study of the 'rules of the game' of the commons as institution. For the basic Ostromian approach, there are many different elements of this rule-bound governance structure that researchers should study. These include the membership of the community, including any membership criteria, the roles within the community occupied by those members, and the 'rules-in-use' that govern and constrain members' behaviours within each role.

For the researchers associated with the Workshop on Governing Knowledge Commons, this framework provides a key methodological starting-point for the study of knowledge commons. However, there are several important differences between common pool resource governance institutions and knowledge commons governance institutions, which require modifications to the basic IAD framework before it can be applied to the latter.[3] The following

[2]For a discussion of the connections between Polanyi and Ostrom's accounts of polycentricity in science, and in social organisation more broadly, see Tarko (2015) and Aligica and Tarko (2012).

[3]As research into knowledge commons within the Ostromian framework has continued, scholars have continued to modify the IAD framework in light of empirical findings. For

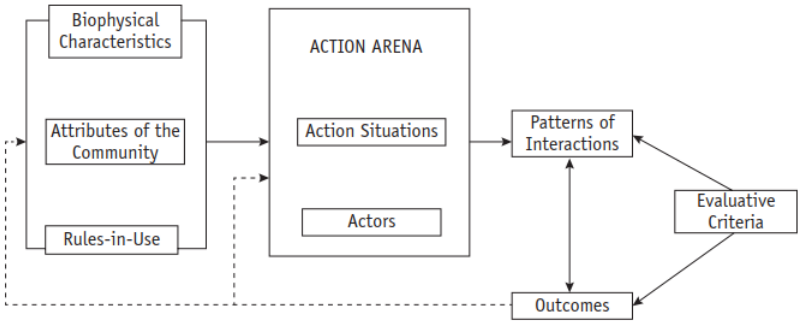


Figure 4.1: The Institutional Analysis and Development framework for studying common pool resource governance communities, as depicted in Strandburg et al. (2017), p. 15, after Ostrom (2005). Licensed under a Creative Commons Open Access licence CC-BY-NC-ND 4.0. To view a copy of this license, visit https://creativecommons.org/licenses/by-nc-nd/4.0.

list, adapted from Strandburg, Frischmann and Madison (2017: 13-14) gives some of the key ways in which knowledge commons governance communities may differ from common pool resource governance communities:

1. knowledge resources must be created before they can be shared;
2. those who participate in knowledge production necessarily borrow from and share with others;
3. because knowledge is nonrivalrous once created, there is often social value in sharing it beyond the bounds of the community that created it;
4. intangible knowledge resources are not naturally defined by boundaries that limit their use;
5. the nonrivalry of knowledge and information resources often rides on top of various rivalrous inputs (such as time or money) and may provide a foundation for various rivalrous outputs (such a money and fame);

this reason, there are small but important differences between the modified IAD framework proposed in Madison, Frischmann and Strandburg (2009), Frischmann, Madison and Sandburg (2014) and Strandburg, Frischmann and Madison (2017). The framework described here is derived from the most recent proposals of the Workshop on Governing Knowledge Commons, as described in Strandburg, Frischmann and Madison (2017).

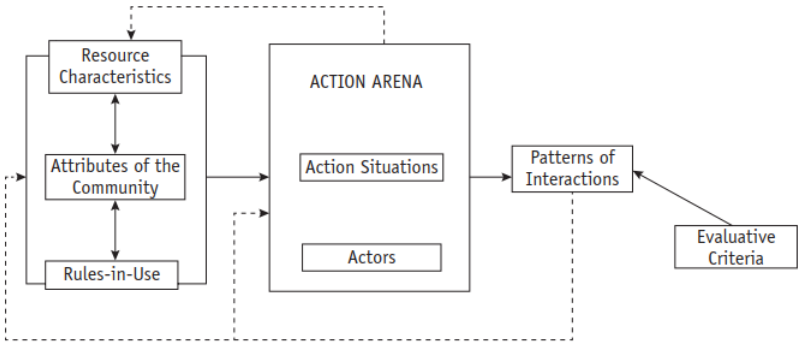


Figure 4.2: The modified Institutional Analysis and Development framework, for studying knowledge commons governance communities, as depicted in Strandburg et al. (2017), p. 15. Licensed under a Creative Commons Open Access licence CC-BY-NC-ND 4.0. To view a copy of this license, visit https://creativecommons.org/licenses/by-nc-nd/4.0.

6. knowledge commons must define and manage not only these resources but also the make-up of the community itself;

7. because of the way in which knowledge resources and communities are co-created, both tend to evolve over time.

Several of these points are captured in the modified diagram of the IAD framework in figure 4.2.

For the purposes of this chapter, I want to draw particular attention to the fifth and third points in this list. The relationship between nonrivalrous knowledge resources and rivalrous outputs is key to the Mertonian model of science that is the core of this thesis: although both knowledge inputs and knowledge outputs are nonrivalrous, a particular form of scientific knowledge output - priority publication - is associated with a rivalrous *reputational* good. Understanding scientific institutions in terms of the way this relationship is managed by institutional rules will be key to the analysis of later chapters.

Moreover, the goal of open inputs advocacy - which I will discuss in more detail in Part III of the thesis - is to expand the scope of broadly available scientific knowledge resources. As the third point in the list above puts it:

> because knowledge is nonrivalrous once created, there is often

> social value in sharing it beyond the bounds of the community that created it. (Strandburg, Frischmann, and Madison 2017: 13)

As I discussed in chapter one, although traditional common pool resources are vulnerable to tragedies of the commons because of their rivalrous character, knowledge resources are not. Therefore one major constraint on the opening and expansion of knowledge commons to users beyond the knowledge-producing community itself is the constraint imposed by *fully socially instituted* scarce resources, such as the reputational dividends associated with priority of publication. I will consider these issues in more depth in Part III of the thesis, when I turn to the discussion of open inputs science.

To sum up the claims of this chapter so far, I have suggested that Polanyi's model of the 'republic of science' helps us understand the *macro*-structure of science as an institution, and that the framework of the Workshop on Governing Knowledge Commons helps us to understand cooperation at the *micro* level within this network structure. In the next section, I will build on the framework proposed by the leaders of the Workshop on Governing Knowledge Commons, by suggesting that the small-scale scientific commons of which the republic of science is composed can often usefully be understood as a specific category of knowledge or cultural commons: *innovation commons* in the sense in which this term is used by Jason Potts (2018).

4.3.1 Innovation commons

The term 'innovation commons' was pioneered by Lawrence Lessig (2002, 2004, 2009). However, in recent years this concept has been developed into a more systematic analytic perspective by a team of researchers at RMIT University, working under the aegis of the 'innovation commons' project.[4] The concept of the innovation commons associated with this work has been most clearly articulated by Potts (2018; cf. Potts 2012, Allen and Potts 2015, 2016), and it is on this work that I will primarily draw in the discussion that follows.

The concept of the innovation commons was developed, by Potts and others (Allen 2017; Davidson and Potts 2016; Potts and Hartley 2015), to fill a gap in the traditional Schumpeterian theory of innovation (Schumpeter 2017, 1994). For this Schumpeterian approach, the dynamics of capitalist

[4]This thesis is a contribution to the RMIT innovation commons project, and is funded by a scholarship associated with that project.

society are driven, in significant part, by a three-stage process: the introduction of an innovation to the economy by an entrepreneur; the transformation of the economy through the 'creative destruction' created by that innovation; and the stabilisation of the economy in a new equilibrium incorporating the innovation. This Schumpeterean analysis is, however, silent on the origin of the disruptive innovation, prior to its introduction to the market by the entrepreneur: innovation itself is regarded as an exogenous factor, in the Schumpeterian framework.

The concept of the innovation commons aims to illuminate the 'zeroth phase' of the innovation process - the development of innovation *prior* to the introduction of that innovation to the marketplace by the entrepreneur (Potts 2018: 1027). For Potts, and others working within the innovation commons paradigm, innovations are often developed within an Ostromian commons of pooled material and knowledge resources, prior to the entrepreneurial stage of the Schumpeterean innovation cycle. Such innovation commons assist the innovation process both in the development of the innovation itself, and in the development of strategies for the introduction of an innovation to the broader market economy. In Potts' words:

> An innovation commons is a species of *knowledge commons* (Hess and Ostrom 2007; Madison, Frischmann, and Strandburg 2009a; Madison, Strandburg, and Frischmann 2016) to minimise the transactions costs of discovering entrepreneurial opportunities about the value of particular positions, resources, information and knowledge that each individual holds. (Potts 2018: 1031)

Put differently:

> An innovation commons is an institution to pool distributed information to discover opportunities that would otherwise be, for an individual, prohibitively costly or inefficient to discover. (Potts 2018: 1030)

Innovation institutions that can be classified as innovation commons include the Homebrew Computer Club, many forms of open source software development, hackerspaces, and blockchain communities (2018: 1030-31). However, the concept of the innovation commons can also be applied to *scientific* commons governance communities. Potts suggests that the 'republic

of letters' of early modern Europe - the "virtual community of scholars formed in 1680-1720 across Western Europe that begat the Industrial Enlightenment and the industrial revolution" - was one early example of a scientific innovation commons (Potts 2018: 1030; cf. Mokyr 2016):

> the *Respublica Literaria* was not just the communications of a trans-European bunch of gentlemen scholars. This was a self-governing community, a group of people who arrived at an effective institutional solution to a fundamental economic problem in the production of knowledge, namely: new knowledge benefits from the co-production of complementary other knowledge. (Potts 2018: 1030)

This thesis applies this concept of the innovation commons to the analysis of scientific production. In the remainder of this section I will elaborate on how the concept of the innovation commons can be applied to the analysis of 'the republic of science', and how the categories of innovation commons scholarship should be understood in this specific institutional context.

4.3.2 Scientific innovation commons

There are two broad ways in which the concept of the innovation commons can be applied to the study of scientific production - each of which represents a different 'perspective' on the scientific innovation process. First, the innovation commons perspective can be integrated into a 'weak' version of the linear model of innovation, as discussed in chapter one (cf. Balconi, Brusoni, and Orsenigo 2010). From this perspective, the innovation process advances in stages, with scientific research as either an originary stage that precedes the others (Bush 1945); a resource available for other key stages of the innovation process (Kline and others 1985); or both. Explaining how scientific innovation operates can therefore provide an account of the 'zeroth stage' of the innovation process. This perspective on innovation commons integrates the category of the scientific innovation commons into a broader picture of the innovation ecosystem, and can be seen as a contribution to the 'national systems of innovation' approach.

I will return to this perspective on scientific innovation commons in Part III of the thesis. However, in the remainder of this section, I will focus on a second perspective on the concept of the scientific innovation commons.

This is a perspective that focuses entirely on the *internal* dynamics of the scientific research process. From this perspective, the scientific innovation commons is one subcomponent of the broader scientific community, and the scientific innovation commons can be seen as playing a role within the *epistemic* economy of the scientific ecosystem.

From this second perspective, scientific innovation commons should be understood as generating contributions to the metaphorical 'marketplace of ideas' of the broader scientific public sphere. Just as the individual scientific innovator proposes ideas that are debated, tested, and evaluated within the broader scientific discourse, so scientific innovation commons can be a 'zeroth stage' in the evolutionary epistemological dynamics of scientific conjectures and refutations. Scientists can collaborate within a scientific innovation commons in a manner that pools intellectual resources, the better to facilitate scientific innovation - and this innovation can then be seen as one contribution to the broader evolutionary dynamics of the 'republic of science'.

If we think of scientific innovation commons in this way, then the moment of 'entrepreneurial innovation' within the scientific research process can be understood as the moment in which a scientific finding is published. Moreover, in this picture, the 'marketplace of ideas' - that is to say, the public domain of scientific findings - is the location of the Schumpeterian or Popperian dynamic of competition between rival claims, rebuttals, conjectures, and refutations. This is the understanding of scientific knowledge discussed in chapter two, under the heading of 'evolutionary social epistemology'. In this analogy, the scientific innovation commons should be understood as the site of scientific production in which material and intellectual resources are pooled by researchers working together to collectively generate knowledge within the scientific epistemic ecosystem. This understanding of the scientific innovation commons connects the concept to the Ostromian case studies discussed above.

With this understanding of the scientific innovation commons in mind, how can Potts' analysis of the innovation commons be applied to this process of scientific innovation *within* the republic of science?

For Potts, there are three defining characteristics of the innovation commons.

> First, there are two distinct resources in the innovation commons - i.e. it is actually *two commons*. One resource is the materials and technology, such as kit, tools, samples, and other pooled and

> shared things.... But the other distinct resource that pools in an innovation commons is information that is relevant to discovering the entrepreneurial opportunity. (Potts 2018: 1036)

This element of the innovation commons framework can readily be applied to the analysis of scientific commons, which frequently pool both the material and the intellectual resources of community members. In the words of an early historian of the Royal Society:[5]

> It was necessary there should be many Heads and many Hands, and Those formed into an Assembly, that might intercommunicate their Tryals and Observations, that might joyntly work, and joyntly consider: that so the improvable and luciferous Phaenomena, that lie scatter'd up and down in the vast Champaign of Nature, might be aggregated and brought into a common store. (Glanvill 1668: 87-8)

The second characteristic of innovation commons is the institution's relationship to *rival* economic governance institutions. In Potts' words:

> the innovation commons can function for *defensive purposes* to protect a technology from capture by alternative institutional forms, and thus from being controlled or monopolised by any one group.... Placing an idea in the commons, or a credible threat of doing so, can be an effective defensive gambit to foreclose alternative institutional modes of development of the new technology or idea (Potts 2018: 1037)

This attribute of innovation commons is relevant to Mertonian science, one of the major historical institution-design principles of which is to preserve scientific knowledge from capture by governance structures that would foreclose 'open findings' - the placing of scientific findings in the public domain. This goal is achieved by Mertonian science in two ways. First by creating an Ostromian commons in which scientific research can take place, outside the immediate control of state or market governance structures. Second, by

[5]For a discussion of the internal structure of the early Royal Society see (Purver 2013; Hunter 1981, 1989, 1994). In future work I hope to discuss in more detail the development and rules of the early Royal Society through the lens of the innovation commons framework.

incentivising the creation of a public domain resource of scientific findings, replacing the institutional incentive structures of intellectual property rights and state or industrial secrets with the incentive structure of 'reputational property rights' (Eamon 1985).

Both of these mechanisms can be seen as 'defensive gambits' that limit control of scientific knowledge by state, corporate, or other powerful actors. This dynamic can be seen in Workshop in Governing Knowledge Commons case studies like the human genome project (Contreras 2014) where an institutional structure for human genome research was established that sought to prevent the 'enclosure' of this prospective, but not yet created, knowledge. The innovation commons framework suggests that this is just one example of a broader function of innovation commons within the scientific ecosystem.

Potts' third characteristic of innovation commons relates to the institution's lifespan:

> ... innovation commons tend to be *temporary* and can rapidly collapse (Allen and Potts 2016).... The collapse occurs when an innovation commons has completed its function by revealing the entrepreneurial opportunity, reducing uncertainty to a level tolerable for entrepreneurial action. At this point the value of the commons to those in it collapses, so they exit and subsequently enter the institutions of the Schumpeterian trajectory, namely innovating firms in competitive markets... The collapse of the innovation commons is not the result of a tragedy, but rather is evidence of its success in producing the common pool resource that those who contributed to the commons sought, namely reduced entrepreneurial uncertainty. (Potts 2018: 1037-8)

This collapse of the innovation commons is a particularly important feature of the scientific innovation commons - and, I will argue, relates to the n-player prisoner's dilemma structure of the scientific innovation commons.

As I have already discussed, the reputational incentive structure of Mertonian science incentivises priority of publication. This means that as a cooperative group gets closer to the creation of the entrepreneurial opportunity - in the case of the scientific innovation commons, the development of a publishable scientific finding - the incentives for the commons to collapse and individual scientists to make an 'entrepreneurial' bid for reputational ownership of the commons' collective output increases. This helps explain

why the history of science contains many cases of contestations over scientific reputational ownership between collaborators (cf. Merton 1961). Of course, an innovation commons need not collapse in this way - in particular, a commons that is persistently generative of an ongoing flow of scientific outputs is less likely to collapse, as the iterated game will continue to reward participants with reputational dividends if the commons persists. Nevertheless, the tendency of scientific innovation commons towards collapse, as their original function of knowledge production is fulfilled, is one of the elements of this category of institution that can inform their categorisation as *innovation* commons.

I will discuss the dynamics of this system in more detail in chapter five. Here my goal is to outline the basic qualitative structure of the model of science I am proposing. In the next section I therefore summarise the qualitative picture of the rules of the scientific game that I have developed so far, and discuss its implications for how we understand the category of the 'scientific commons'

4.4 The internal structure of Mertonian science

In the chapter so far, I have outlined the key features of the qualitative model of science I am proposing. This model has two key components, operating at different scales of analysis. At the *macro* level, this qualitative model adopts Polanyi's account of 'the republic of science'. Following Polanyi, I see science as a global network of researchers composed of many smaller, nested, partially overlapping subcommunities. As with Adam Smith's concept of the 'invisible hand', no individual decision-maker is responsible for the epistemic outcomes of the system as a whole, but many individuals pursuing their own research produce an emergent body of consensus knowledge that we can regard as the scientific community's overall output.

At a *micro* level, I propose using Ostromian resources to study the 'rules of the game' that structure small-scale scientific subcommunities. In section 4.3, I discussed the Workshop on Governing Knowledge Commons, and its extension of Ostrom's 'Institutional Analysis and Development' to the study of scientific knowledge commons. I then proposed adopting a specific theoretical perspective on these small-scale scientific knowledge commons: the framework of the 'innovation commons' proposed by Potts and others.

Scientific innovation commons, I argued, are small-scale, local, tempo-

rary, self-governing research communities formed to explore a specific scientific problem space. These institutions exist in the moment of the scientific innovation process when the character of any scientific innovation is unclear, and they often collapse once a scientific discovery has been achieved. An innovation commons allows scientists to pool materials both physical and intellectual, in order to leverage cooperative dynamics into achieving research outputs that would be impossible or difficult for a single individual to achieve. Innovation commons are not the *only* micro-level locus of scientific innovation - scientists may, for example, engage in individual scientific innovation which does not rely on pooled resources within an Ostromian commons, or participate in larger scale scientific communities that do not exhibit the qualitative attributes of an innovation commons. However I suggest that scientific innovation commons are a significant component of the scientific innovation ecosystem, and that the theory of the scientific innovation commons provides a novel lens onto categories of social organisation within the scientific research process that have traditionally been captured using the concepts of 'communities of practice' (Wenger 1998) or 'invisible colleges' (Crane 1972). Specifically, the concept of the Ostromian innovation commons illuminates the nature of the "small cooperative 'networks' of information-sharing" discussed by Dasgupta and David in their discussion of the new economics of science (Dasgupta and David 1994: 504).

Putting these claims together with the understanding of scientific institutions developed in Part I of the thesis, my qualitative picture of science as an institution therefore has several distinct elements:

- at the macro level, this thesis sees science as complex network of partially overlapping subcommunities;
- this overall community is structured by the Mertonian norms of communism, universalism, disinterestedness, and organised scepticism;
- moreover, the reputational incentive structure associated with priority of publication is a key determinant of scientists' behaviour within this community;
- at smaller scales, science is a nested structure of Ostromian subcommunities, each with its own specific norms, associated with distinct disciplines, subdisciplines, sectors, organisations etc.;

- one key category of small-scale institution within this nested structure is the scientific innovation commons, within which scientists pool resources in order to engage in higher levels of scientific innovation than would be possible individually.

The picture presented in this chapter therefore elaborates the core Dasgupta and David model in several ways. Where Dasgupta and David's model of science as n-player prisoner's dilemma captures a key structure of science as institution, Dasgupta and David do not analyse in depth the larger scientific institutional structures within which prisoner's dilemmas may be played. Differentiating between different subcomponents of this structure using Ostromian analytic tools will allow me, in subsequent chapters, to give a more nuanced account of the strategic terrain that confronts working scientists as they attempt to navigate the reputational economy instituted by Mertonian 'rules of the game'. In later chapters I will explore this issue by developing a series of formal, game-theoretic models of working scientists' strategic decision-making, and its implications, within a nested Mertonian 'republic of science'. Before I turn to the development of these formal models, however, I want to discuss the implications that this qualitative picture of science as an institution has for our understanding of the category of scientific commons.

4.4.1 Three categories of scientific commons

In section 4.3, I discussed the role of Ostromian commons within the republic of science. In this section, I will connect this discussion to other typologies of science as commons, in order to highlight the overlaps and differences between the framework I am adopting in this thesis, and other similar frameworks within the prior science studies literature. This will enable me to distinguish between three broad categories of scientific commons, in a way that will have implications for the discussion of 'open inputs commons', in later chapters.

In his work on knowledge commons, Peter Levine has distinguished between what he calls "a libertarian commons" and "an associational commons" (Levine 2007: 250). In Levine's words:

> In a *libertarian commons*, anyone has a right to use (and sometimes also to contribute to) some public resource.... In contrast, an *associational commons* exists when some good is controlled by a group. (Levine 2007: 250-1)

Thus for example knowledge resources published on the open web belong to a libertarian commons, which can be accessed at low cost by anyone with the appropriate technology. Knowledge resources shared within a closed research group, by contrast, belong to an associational commons. For Levine, associational commons have the benefit that, because the commons is controlled and managed by a closed group,

> it can litigate and lobby to protect the public good of which it is the steward. (Levine 2007: 251)

Levine is not the only scholar to propose this distinction, which aligns with James Boyle's similar distinction between a commons and the public domain. For Boyle, the public domain is

> the "outside" of the intellectual property system, the material that is free for all to use and build upon. (Boyle 2003a: 1)

This category of the public domain is thus defined by its accessibility, in opposition to "enclosure" of knowledge by intellectual property rights. Boyle contrasts this category of general accessibility with the category of the commons. In a commons there are no *individual* property rights, but there is *collective* control of intellectual resources, which may greatly restrict access to those resources for those outside the commons community. Thus while both the public domain and the commons can be seen as the "opposite" of enclosed private property, they are "opposites" in different senses (Boyle 2003a: 8,b).

These distinctions - between libertarian and associational commons, or between public domain and commons - provide important resources for thinking about scientific knowledge-sharing. However, in the chapters that follow I will use my own tripartite distinction, between "open", "closed", and "Ostromian" knowledge commons, which I argue captures the issues relevant to the focus of this thesis more precisely. I will use "open commons" to refer to knowledge resources that - like Boyle's public domain and Levine's libertarian commons - have low barriers to access and limited rules governing how their knowledge resources are used by those who access them. I will use "closed commons" to refer to knowledge resources that are only accessible to members of a relevant commons governance community. Finally, I will use "Ostromian commons" to refer to the governance community itself, and to the rules and practices that structure the behaviour of this community.

This tripartite distinction has several advantages over Levine and Boyle's bipartite distinctions. In particular, it distinguishes between the commons community itself, and the status of the knowledge that the community shares. An Ostromian commons governance community determines the rules of the game that govern knowledge sharing. Those rules of the game may, however, result in different types of knowledge sharing: an Ostromian commons may choose to share knowledge resources only among Ostromian commons members, resulting in a closed knowledge commons, or it may choose to share its knowledge resources more broadly, including with people outside the Ostromian commons itself, resulting in an open knowledge commons.

From this perspective, both open and closed knowledge commons can be produced by the rules of the game instituted by Ostromian commons. It is therefore important to see (in Levine's terms) 'associational' and 'libertarian' commons as partly orthogonal categories, rather than as direct opposites.

Although I will use these terms throughout the remainder of the thesis, it is important to acknowledge that these are 'ideal typical' categories. As many theorists have emphasised (e.g. Callon 1994), no open knowledge commons is fully open: even a knowledge commons with very low barriers to access has *some* barriers to access. For example, scientific publications available for free on the open web are still difficult to access for those who do not have internet access, do not read the language in which the publications are written, etc.

For this reason Buchanan has proposed that the category of public good is misleading, and is better reconceptualised as a club good with a very large club (Buchanan 1965). Adopting this insight, we can think of the distinction between a closed knowledge commons and an open knowledge commons as one of degree: these categories represent poles of a spectrum rather than a sharp binary.

For the purposes of this thesis, the more relevant subcategories of commons are those most naturally grouped under the heading of Ostromian 'closed' commons. Here it is possible to distinguish between the republic of science as a whole, different disciplinary communities, different membership organisations such as professional societies, universities, etc., and different innovation commons in the sense defined in section 4.3.1. All of these communities can be understood as commons in some sense, in that they are membership clubs participating in the collective management of knowledge resources. But these different commons have different rules and govern different subsets of knowledge resources.

One more nuance needs to be discussed before I move on - this is the rela-

tionship between knowledge and Ostromian commons, in the senses in which I am using these terms. One of the key explanatory elements of this thesis is the idea that open *and* closed knowledge commons are instituted by the rules of the game of Ostromian scientific commons. Thus the reputational incentive structure of the Ostromian commons of local Mertonian scientific institutions incentivises the publication of scientific findings within an open knowledge commons that is also produced by this incentive structure. That is to say: an *open* knowledge commons is produced by *Ostromian* scientific commons. At the same time, this reputational incentive structure disincentivises the sharing of intermediate inputs within such an open commons, but may incentivise input sharing within a smaller research team structured as an Ostromian innovation commons. The rules of the Mertonian game therefore contribute to the structure of Ostromian commons which produces *both* open *and* closed knowledge commons.

This point will be particularly relevant in Part III of the thesis, when I turn to the relationship between traditional 'Mertonian' science and open inputs science. For now, though, the goal of the remainder of Part II of the thesis is to develop a series of formal agent-based models based on the qualitative picture of the structure of science as an institution I have elaborated above.

Chapter 5

Modelling the republic of science

5.1 Introduction

In chapter four I discussed the qualitative picture of science as an institution that informs this thesis. That picture built on Dasgupta and David's model of science as an n-player prisoner's dilemma structured by reputational incentives to priority of publication. It extended that Dasgupta and David model in three key ways. First, I embedded Dasgupta and David's concept of Mertonian science as n-player prisoner's dilemma within a Polanyian understanding of the 'republic of science'. Second, I drew on the approach of the Workshop on Governing Knowledge Commons to analyse the 'action arena' of science as an institution in Ostromian terms. Third, I suggested that one of the key local institutions of science understood in this way is the 'innovation commons', in the sense in which that term is used by Jason Potts. This framework then allowed me to distinguish between three different categories of scientific commons, in a way that will be analytically useful in later chapters of the thesis.

This framework provides the basic building blocks for the models of scientific institutions developed in this thesis. In the remainder of Part II, I will use this framework to develop formal agent-based models of Mertonian science. I will do this in relation to the two core issues of open inputs that motivate the thesis: the practice of cooperative input-sharing by working scientists, and the epistemic impact of that practice on knowledge outputs.

In this chapter I begin this process, by focusing on the cooperative dimension of scientific practice. To foreshadow the argument of the chapter: I will

begin by reviewing in more depth Dasgupta and David's model of science as n-player prisoner's dilemma. In section 5.2 I present the formal structure of an n-player prisoner's dilemma, and discuss some of the ways in which the dilemma can potentially be 'solved'. In particular, I discuss the 'classical' solution of conditionally cooperative strategies adopted by game players.

In section 5.3 I argue that Ostrom's concept of a commons governance community - and Potts' related concept of an innovation commons - can be interpreted in terms of the deployment of conditionally cooperative strategies to 'solve' the prisoner's dilemma. I argue in particular that the micro-level institution of the scientific innovation commons, which I discussed in chapter four, can be understood in terms of such conditionally cooperative strategies.

In section 5.4 I then turn to the macro structure of the scientific commons, and the relation between the innovation commons subcommunities discussed in chapter four. I argue that Polanyi's 'republic of science' serves as a social environment within which social selection of community norms can operate. I review the principles of evolutionary game theory and, in sections 5.6 and 5.7, I apply this approach to the analysis of the 'republic of science'. Finally, in section 5.8 I discuss the evolutionary dynamics of group selection among innovation commons described in this formal model.

5.2 Cooperative outcomes in the n-player prisoner's dilemma

In chapter two, I discussed Dasgupta and David's game-theoretic formalisation of Merton's model of scientific incentive structures. For Dasgupta and David, scientists' incentives for sharing the intermediate inputs of scientific production have the structure of a prisoner's dilemma game. In its two player form, this prisoner's dilemma can be represented as follows:

	C	D
C	$a-c, a-c$	$-c, a$
D	$a, -c$	$0, 0$

Where $a > c > 0$

In chapters two and three, I discussed possible equilibrium outcomes for this game. In its one-round version, there is only one equilibrium - uni-

versal defection; in repeated versions of the game cooperative outcomes are possible, via conditionally cooperative strategies. However, these results are only partly relevant to the problem of scientific input-sharing. As Dasgupta and David observe, most real world scientists are engaged, not in iterated two-player games, but in iterated n-player games. In Part I of the thesis, I postponed the discussion of such n-player games; in this section I return to that topic. First I discuss how to formally represent an n-player prisoner's dilemma; then I discuss possible cooperative equilibria in the iterated version of the game; finally I describe the contribution evolutionary game theory can make to the analysis of this game form.

5.2.1 The iterated n-player prisoner's dilemma

The first problem analysis of the n-player dilemma must confront is how to generalise the two-player game to an n-player scenario. In his influential *Micromotives and Macrobehavior*, Schelling argues that an n-player prisoner's dilemma - or, in Schelling's vocabulary, a uniform multi-person prisoner's dilemma - has the following features:

> 1. There are n people, each with the same binary choice and the same payoffs.
> 2. Each has a preferred choice whatever the others do; and the same choice is preferred by everybody.
> 3. Whichever choice a person makes, he or she is better off, the more there are among the others who choose their unpreferred alternative.
> 4. There is some number, k, greater than 1, such that if individuals numbering k or more choose their unpreferred alternative and the rest do not, those who do are better off than if they had all chosen their preferred alternatives, but if they number less than k this is not true. (Schelling 1978: 218)

In *The possibility of cooperation* Michael Taylor argues that this definition of an n-player prisoner's dilemma is too restrictive. The n-player dilemma, as Taylor puts it, assumes:

> that a player's payoffs in each constituent game depend upon two things only: his strategy (C or D) in that constituent game, and *the number of other players choosing* C in that game. (Taylor 1987: 83)

N-player games can therefore be modelled with two functions; Taylor denotes these functions as $f(v)$ and $g(v)$ - I denote them as $V(C|i)$ and $V(D|i)$. In this notation, i denotes the number of *other* players who play cooperatively; given i, $V(C|i)$ denotes the payoff to a player who chooses strategy C; $V(D|i)$ denotes the payoff to a player who chooses strategy D.

Taylor defines an n-player prisoner's dilemma in the following way:

(i) $V(D|i) > V(C|i)$ for each value of $i \geq 0$
(ii) $V(C|N-1) > V(D|0)$

For Taylor these two principles alone are both necessary and sufficient to define a game as an n-player prisoner's dilemma. Taylor also stipulates an additional principle:

(iiia) $V(D|i) > V(D|0)$ for all $i > 0$

though Taylor notes that in practice a stronger version of (iii) often holds:

(iiib) both $V(C|i)$ and $V(D|i)$ are monotonically increasing with i

Working with this last definition we can - again following Schelling - represent any game visually by graphing these two functions. Figure 5.1 represents a range of possible n-player games that meet criteria (i) (ii) and (iiib).

Again following Schelling, these graphs all arbitrarily assume $V(D|0) = 0$. Therefore the i at which $V(C|i) = 0$ gives us the minimum i at which a viable coalition of cooperators can be formed. For any i such that $V(C|i) > 0$ the members of this coalition (of size $i + 1$) would all have higher payoffs than they would in the coalition's absence.

Of course, what makes this game a prisoner's dilemma is that such a coalition is *not* stable in a one-round version of the game. Since $V(D|i) > V(C|i)$ for each value of $i \geq 0$, cooperation is always a strictly dominated strategy in a one-round game, even when such a coalition is in principle possible. Just as with the two-player dilemma, however, this issue can potentially be resolved by shifting to an iterated version of the game.

In chapter four of *The possibility of cooperation*, Taylor explores the conditions under which an iterated n-player dilemma can result in cooperative outcomes. He begins by considering the scenario in which $k = n$; that is

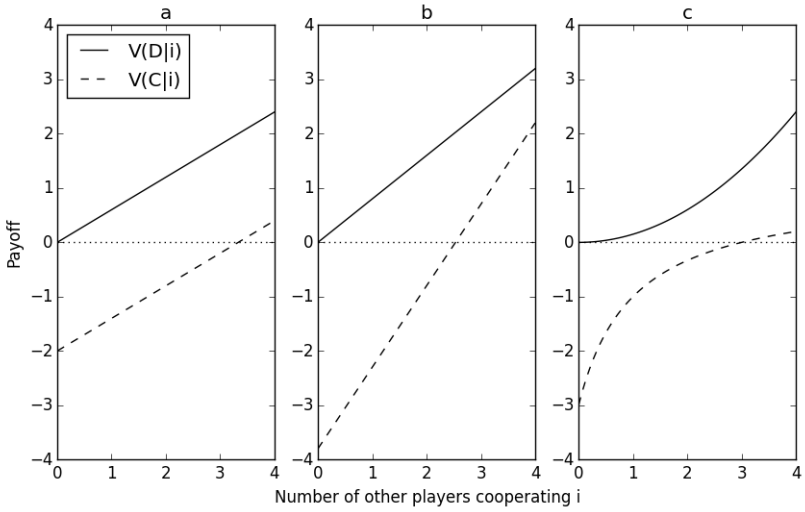


Figure 5.1: Examples of n-player prisoner's dilemmas

to say, in which the size of the minimal viable coalition is the same as the group size. Here, if any player defects this will result in payoffs to all players below those available from universal defection. Provided discounting is low enough, all players therefore have an incentive to play the 'grim trigger', or some other conditionally cooperative strategy, potentially resulting in a universally cooperative equilibrium.

In scenarios in which $k < n$, however, the situation is more complex. Here *some* players can defect without eliminating the incentive for the remaining players to play cooperatively. Given the structure of the dilemma, it is in the interest of every player to be one of the defectors, leaving others to mutually enforce the conditionally cooperative coalition. An n-player prisoner's dilemma of this kind therefore in practice takes on the form of an n-player game of chicken, in which the players compete to credibly commit themselves to noncooperation, in the hope that the other players can thereby be induced to play cooperatively. In a game of chicken *any* outcome is possible, and therefore the game-theoretic apparatus here provides no predictive purchase.

In Taylor's words:

> Speaking informally (which is all we can do here), it is pretty clear that Cooperation amongst a relatively large number of players is

> 'less likely' to occur than Cooperation among a small number. For a start, the more players there are, the greater is the number of conditions that have to be satisfied - the conditions specifying that the right kind of conditionally Cooperative strategies are present and those specifying the inequalities that all the Cooperators' discount rates must satisfy. But the main reason for this new 'size' effect is that Cooperation can be sustained only if conditional Cooperators are present and conditional Cooperators must be able to monitor the behaviour of others. Clearly, this monitoring becomes increasingly difficult as the size of the group increases. (105)

This argument is plausible - but, as Taylor notes, it is informal, relying on principles that fall outside the formal game-theoretic analysis that is the focus of his book. In the next section I will therefore look at how this argument relates to qualitative Ostromian analysis of scientific innovation commons.

5.3 Conditionally cooperative strategies and the Ostromian commons

In the previous section I reviewed the structure of the n-player prisoner's dilemma. I discussed the possibility of conditionally cooperative strategies resulting in mutually beneficial cooperative play within the dilemma, as well as the difficulty of forecasting such conditionally cooperative equilibria. In this section I will connect these conditionally cooperative strategies to the Ostromian apparatus I discussed in chapters three and four.

Ostrom's own work often contrasts her approach to that of the traditional prisoner's dilemma analysis of commons dynamics. In *Governing the commons*, for example, Ostrom notes the grip that the metaphor of the prisoner's dilemma has on the imagination of some researchers and policy-makers, leading them to overgeneralise the results of the simple prisoner's dilemma to an excessively wide range of policy scenarios (Ostrom 1990: 3-8). As Ostrom also argues, however, the flaw in such inferential reasoning does not necessarily reside in the structure of the prisoner's dilemma model itself:

> The results from more recent work, particularly work focusing on the dynamic aspects of relevant empirical settings, have begun

> to generate more optimistic predictions than did earlier models (Ostrom 1990: 7)

Thus while one response to the shortcomings of the traditional prisoner's dilemma model is to abandon it in favour of alternative institutional analyses and frameworks, another is to focus on the emergence of cooperation *within* the constraints of the dilemma.

This approach can also be applied to the interpretation of Ostrom's own work. Are Ostromian commons governance institutions best understood as *alternatives* to the institutional framework of the prisoner's dilemma, modifying the overarching rules of the economic game to achieve more cooperative outcomes, or are they best understood as *solutions* to the prisoner's dilemma, achieved within the strategic space of the dilemma itself? Some researchers (see for example Fennell 2011: 14) have adopted the former interpretation of Ostrom's work, building on Ostrom's own criticism of the privileging of the prisoner's dilemma (e.g. Ostrom 1990: 29-30).

In this thesis, however, I take the second approach. That is to say, I assume that the Ostromian governance solutions to the prisoner's dilemma problems I am studying can usefully be understood as solutions to the prisoner's dilemma achieved *within* the strategic space of the dilemma itself. Without making general claims about Ostromian cooperation as such, I adopt the analytic strategy of treating Ostromian scientific commons as patterns of cooperative practice that operate within an institutional framework that can usefully be modelled as an n-player prisoner's dilemma.

From this perspective, the emergence of Ostromian solutions to the problem of cooperation within the prisoner's dilemma can be interpreted as game players adopting conditionally cooperative strategies that institute norms of practice within the dilemma's strategic space. As Ostrom emphasises, sanctions - specifically gradated sanctions - are a key component of effective commons governance institutions. These sanctions, I suggest, can usefully be understood in terms of conditionally cooperative strategies adopted by players of the game. In terms of the macro-/micro-social categories I discussed in the previous chapter, science as prisoner's dilemma can be seen as a product of the overarching structure of Mertonian institutions, while the norms and practices of Ostromian scientific innovation commons are instituted by agents as local solutions to the collective action problems generated by that overarching structure.

Empirical studies of specific Ostromian scientific commons, like those car-

ried out by The Workshop on Governing Knowledge Commons, are focused on the specific rules instituted by these local governance communities. The focus of this thesis, however, is not on the structure of any given innovation commons, but on the dynamics of the system as a whole. To study this, it is necessary to shift from the study of individual innovation commons, to the study of the evolutionary dynamics that operate across the entire scientific institutional space.

5.4 Evolutionary game theory

Traditional game theory, of the kind deployed by Schelling and Taylor, presupposes rational, strategic players aiming to maximise their payoffs by considering the rational play of other strategic players operating in comparable social roles - an approach which frequently leads to a very large number of possible equilibria.

One major strand of evolutionary game theory aims to address this problem by transforming the way in which strategic play is understood. As I discussed in chapter three, in such evolutionary game theory, players are not rational strategic agents choosing from the entire universe of possible strategies; rather, play takes place in a large *population* of players, among whom strategies are distributed. Players interact, receiving payoffs, and then successful strategies are selected for by the replication of high payoff strategies, and the elimination of low payoff strategies. In this way, high payoff strategies tend to spread across the population without the requirement for any rational foresight or planning by the players of the games.

As I discussed in chapter three, evolutionary game theory was pioneered by evolutionary biologists (Lewontin 1961; Smith 1972), and introduced into social science as a resource for studying the dynamics of social behaviour (Axelrod and Hamilton 1981; Axelrod 2006; Sugden 1986). Biologists making use of evolutionary game theory typically take the components of Darwinian natural selection as their models' referents: strategies are interpreted as phenotypes, players as organisms, and payoffs as an index of 'fitness'. For economists making use of these resources, by contrast, strategies can be interpreted in traditional social and behavioural terms. For economists using this framework, the players are interpreted as *boundedly rational*, and the replication mechanism is a *learning process* in which other players are observed and successful strategies adopted. This evolutionary approach therefore ad-

dresses the common objection to traditional game theory that its concept of rationality is hyperbolic. Moreover, this approach can accommodate stochastic processes, path dependency, and other ways in which 'history matters' for social change.

This chapter takes this agent-based modelling approach. In particular, it draws on the growing literature that uses evolutionary agent-based modelling to study the dynamics of n-player prisoner's dilemma games (Gotts, Polhill, and Law 2003; O'Riordan and Sorensen 2008; Rezaei, Kirley, and Pfau 2009). In the next section I adapt a specific model from this literature (O'Riordan and Sorensen 2008) to represent cooperative intermediate input sharing within Polanyi's republic of science.

5.5 Description of the model

To represent Polanyi's republic of science, I constructed a model in the agent-based programming language NetLogo.[1] The model is a replication of O'Riordan and Sorensen (2008), which I will develop and extend in later sections.

The model consists of a graph-theoretic environment: each agent is a node on a graph, linked to four neighbours by undirected weighted edges. The graph is a structured torus composed of groups of five agents, each group linked to adjacent groups through its outermost members. There are two classes of edge: internal and external. Internal edges link members of a group to other members of the same group - these edges always have a weight of 100. External edges link an outer member of one group with an outer member of another group - these edges are of user-determined variable

[1]The code used in this chapter as well as in section 8.2 below can be downloaded from the NetLogo modelling commons at: `http://modelingcommons.org/browse/one_model/6113`. The simulations data used to generate the graphs in this and subsequent chapters is available for download from: `https://tinyurl.com/ydjl783e`. The NetLogo modelling language itself is available from: `https://ccl.northwestern.edu/netlogo/download.shtml`. The model used in this chapter also makes use of the NW network analysis extension, which is available via GitHub at: `https://github.com/NetLogo/NW-Extension/releases`. Because the in-browser JavaScript version of NetLogo does not include the NW extension, this model cannot be run directly via the NetLogo modelling commons, but must be downloaded from the commons and run using a desktop version of NetLogo that includes the NW extension. The Netlogo software is authored by Uri Wilensky and developed by Northwestern University's Center for Connected Learning and Computer-Based Modeling.

weight, within the range [0,100]. Two strategies are available to the players: cooperate and defect; each agent is randomly assigned a strategy at the start of each simulation, with a user-selected probability of playing cooperatively. All the simulations in this chapter begin with a .5 probability of agents playing cooperate or defect.

In each round of play, every agent creates a game group composed of themselves and a set of their neighbours. Which neighbours will participate in the game is determined stochastically; the probability of their selection is determined by the weight of the edge linking them to the central player. Thus if an edge is of weight 100 they will certainly be selected; if the edge is of weight 0 they will certainly not be selected. In all the simulations described below, internal edges have weight 100 - every agent will therefore always play against their internal neighbours. External edges are variable in weight.

Once groups of players are selected, an n-player prisoner's dilemma is played, of the form described in equations 5.1 and 5.2 below:

$$V(C|i) = (Bi/N) - c \tag{5.1}$$

$$V(D|i) = Bi/N \tag{5.2}$$

Again following O'Riordan and Sorensen, the values for B and c used throughout this chapter are 5 and 3 respectively; these values are selected because they ensure a prisoner's dilemma for all games explored in this chapter. N denotes the number of players in any given game; i the number of *other* players choosing the strategy Cooperate.

On every round of play every agent participates in every game into which they have been selected. Each agent then receives as overall payoff the average of the payoffs received in each individual game.

Once the round of play is complete, every agent can 'learn' from the behaviours of others. There are two learning mechanisms in the model. In the first 'local' mechanism, each agent can learn from its immediate neighbours. The probability of adopting a neighbour's strategy is given by the equation

$$\frac{w(x,y).f(y)}{\sum_{z \in s_adj(x)} w(x,z).f(z)} \tag{5.3}$$

Here $s_adj(x)$ denotes a group of immediate neighbours chosen stochastically by edge weight, $f(y)$ is the average payoff of an agent y in the previous round of play, and $w(x,y)$ is the weight of the edge between agents x and y.

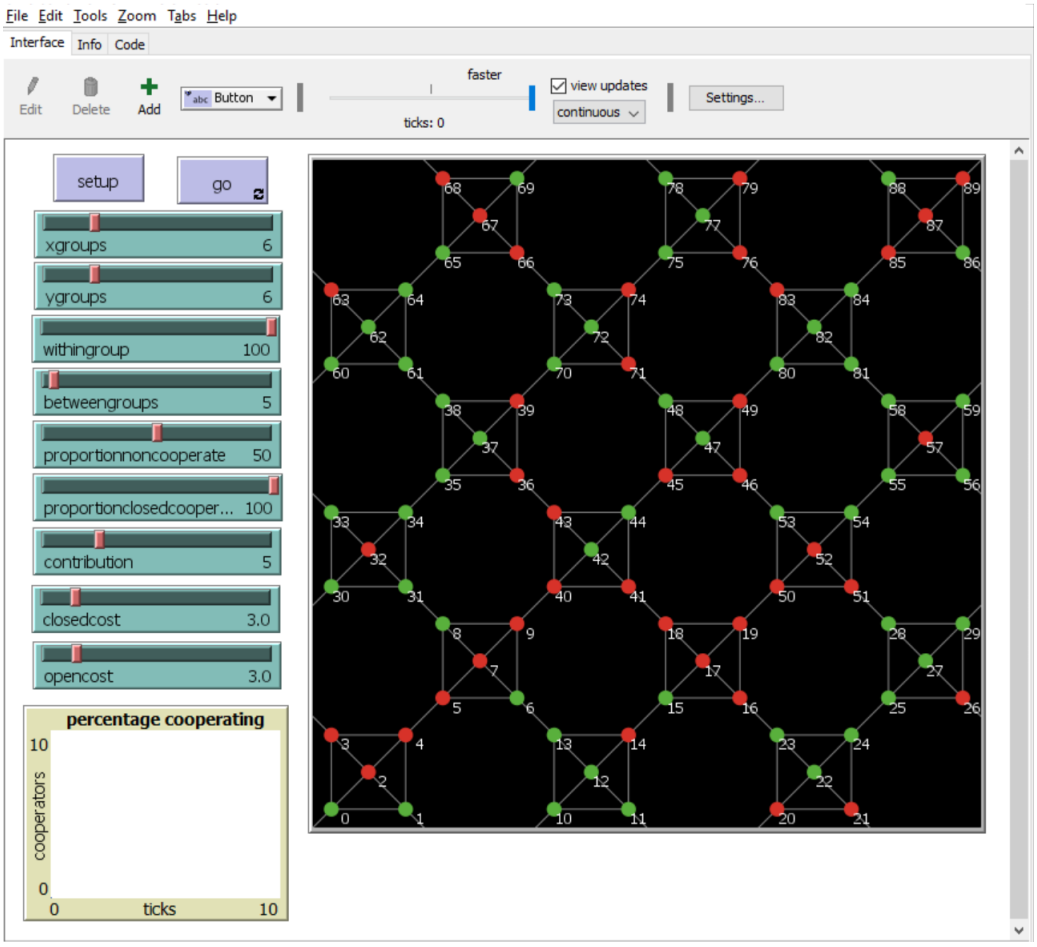


Figure 5.2: The user interface of the NetLogo model, with an illustrative graph of 90 agents.

Every m rounds, every player also has the chance to learn from an extended neighbourhood, encompassing not just their immediate neighbours, but also those neighbours' neighbours. In this learning mechanism the probability of adopting another agent's strategy is given by the equation

$$\frac{w(x,y).f(y)}{\sum_{z \in adj(adj(x))} w(x,z).f(z)} \tag{5.4}$$

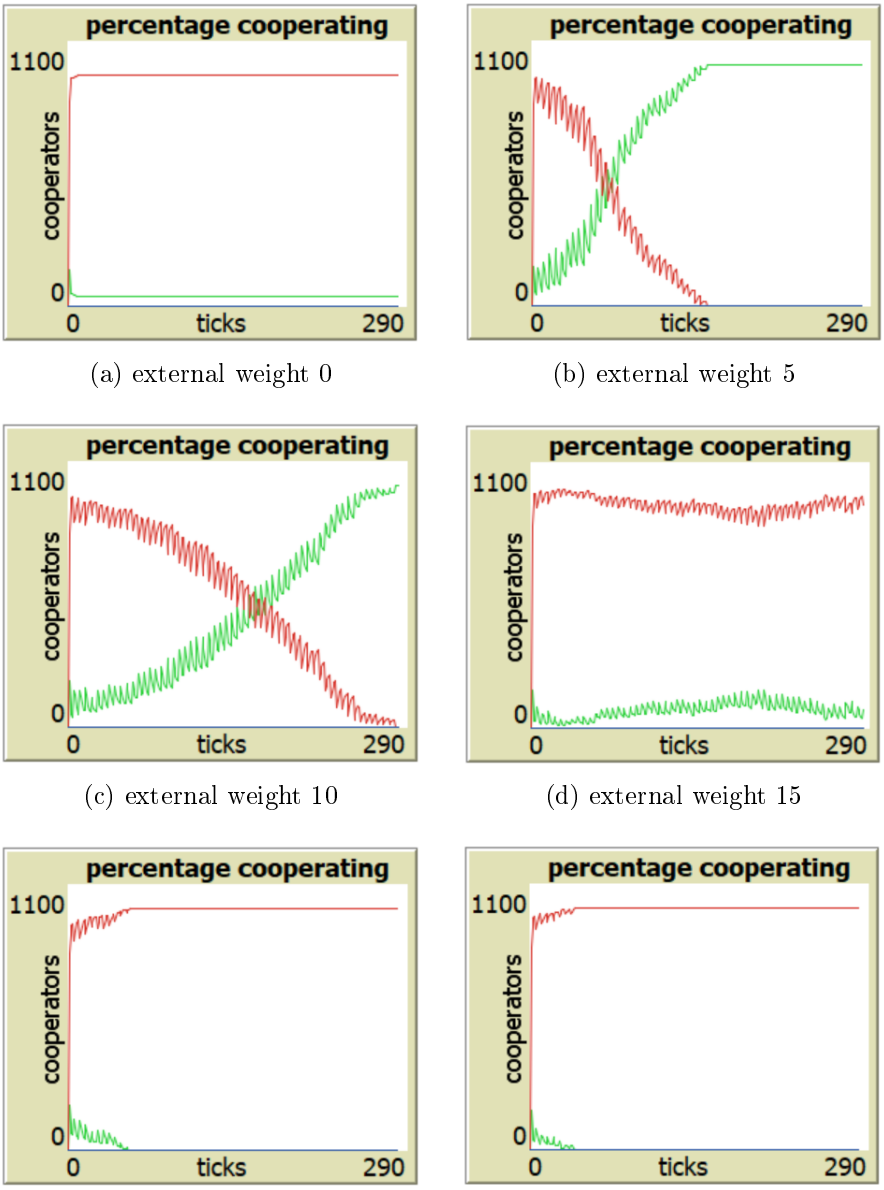
Here $adj(adj(x))$ denotes the entirety of the primary agent's extended neighbourhood - no stochastic group selection takes place. The weight of the path between x and z is the product of the two component paths' weights.

O'Riordan and Sorensen adopt an m of 4 rounds, on the grounds that "four generations is usually sufficient to allow a local community reach an equilibrium point" (162), and I follow them in this practice.

The goal of deploying *two* learning principles or update rules, is to represent the dual relationship of agents to their community. On the one hand agents have a set of local neighbours with whom they regularly interact; on the other hand, agents will also sometimes look outside this immediate community, to evaluate how neighbouring communities are behaving. This parallels Polanyi's account of scientists with an immediate community of shared expertise, and an extended community of *partially* shared expertise: we can understand such agents as engaging in *frequent* collaboration and communication with those in their immediate communities, and *less frequent* collaboration or communication with those in their extended communities.

5.6 Findings

I ran a series of simulations using NetLogo's BehaviorSpace tool, each with 1000 agents, varying the external edge weight from 0 to 100, in increments of 1. For each value of external edge weight, I ran 20 simulations of 300 rounds each. Of these final figures, the majority (96.3%) are stable equilibria with no possibility of further change; the remaining outcomes (3.7%) were still undergoing dynamic change after 300 rounds. Figure 5.3 shows six representative behaviours of the model, for different external weights; each graph shows the number of cooperative and noncooperative agents in the model over the first 280 rounds. Figure 5.4 shows the number of cooperators in round 300, as a mean of 20 simulations for each external weight.



(a) external weight 0

(b) external weight 5

(c) external weight 10

(d) external weight 15

(e) external weight 20

(f) external weight 25

Figure 5.3: Representative simulations. The green lines indicate the number of cooperative agents; the red lines the number of non-cooperative agents.

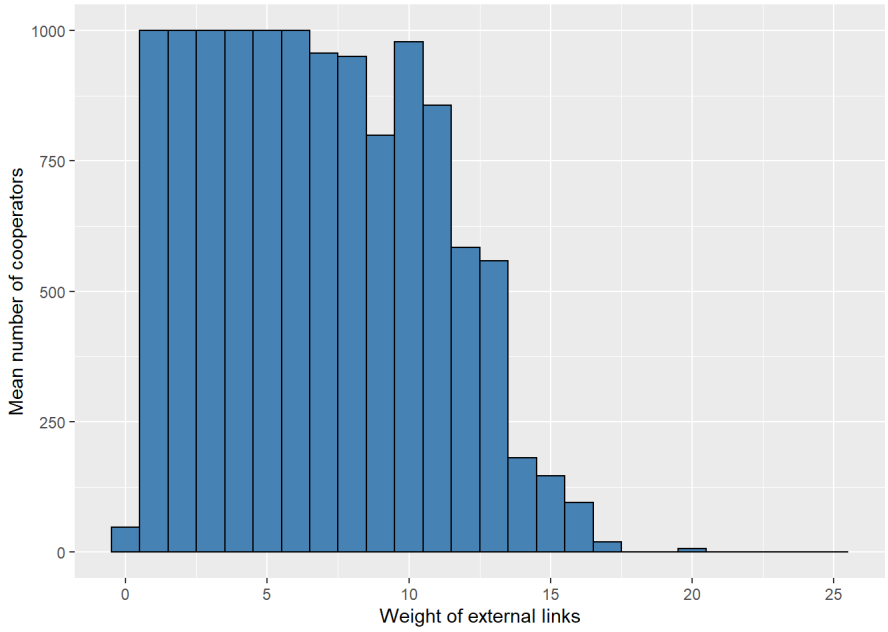


Figure 5.4: Average number of cooperators after 300 rounds of play

In a scenario in which the edge weight between groups is 0, each group is autarkic - no interaction is possible between the groups. In this scenario, each group functions as an independent prisoner's dilemma: if a group contains *any* non-cooperative players, it will move towards the only possible equilibrium of universal defection. If a group contains *all* cooperative players then it will remain in the equilibrium of universal cooperation, because there is no mechanism via which the non-cooperative strategy could enter the group. In a five player group, where each player has a .5 probability of playing cooperatively, the probability of a group achieving a cooperative equilibrium is therefore $.5^5 = .03125$. This is consistent with the average number of cooperators at edge weight 0 across 20 simulations (47.75 of 1000).

By contrast, in a scenario in which the edge weight between groups is 100, we are dealing with a 'flat' or unstructured graph; here universal non-cooperation quickly spreads across the entire graph. The same is true for graphs in which the edge weight between groups is as low as 25. 100% of the graphs with external edge weight 25 or higher resulted in universal defection.

The interesting scenarios, therefore, are those in which the external edge weight is greater than 0 and less than 25.

Figure 5.5 shows the movement from cooperative to noncooperative outcomes as the edge weight is experimentally increased from 0 to 25. As the graph shows, heavily weighted external edges - of 20 or more - reliably produce non-cooperation.

However, with lower external edge weights, cooperative behaviour can spread across the graph. With external edge weights of 6 or less this happened with high reliability. In the range of 6-20, cooperative outcomes and noncooperative outcomes both emerged; however, in general the lower the external edge weights the more likely cooperative behaviour was to spread.

In general, therefore, the simulations suggest that weak external ties between groups with strong internal ties encourage the spread of cooperative behaviour. By contrast, strong ties between groups discourage the spread of cooperative behaviour, and no ties between groups renders the spread between groups of cooperative - and uncooperative - behaviour impossible.

5.7 Discussion

The evolutionary dynamics that generate this result are clear. In a model of groups with weak external ties the two-part update rule permits a *de facto* form of group selection to operate. Because ties between groups are weak, local equilibria can easily emerge within groups. From a starting random distribution of strategies, cooperative equilibria are formed within groups that contain no noncooperative players. In a graph with strong external ties, these local cooperative equilibria are easily 'invaded' by noncooperative strategies present in nearby groups. Weak ties between groups function as a 'barrier' that prevents this very rapid spread of non-cooperative play.

Once local equilibria are established, the second part of the update rule then allows members of groups to extend their sphere of observation, and more easily learn from the members of neighbouring groups. The nature of the prisoner's dilemma is that, within any given game, noncooperative play will always strictly dominate cooperative play, but universal cooperation provides better payoffs than universal defection. A player observing members of their own group is therefore likely to observe noncooperative players receiving higher payoffs than cooperative players. But an agent observing multiple groups is likely to observe cooperative players receiving higher payoffs than noncooperative players, because members of groups in cooperative equilibria will always receive higher payoffs than members of groups in noncooperative

equilibria. The observation of neighbouring groups therefore allows cooperative play to spread - provided that groups are sufficiently 'insulated' by weak external ties that cooperative equilibria can be sustained over the medium term. A form of group selection is operating.

Provided there are strong ties *within* groups, and weak ties *between* groups, scientific subcommunities can serve as locations where cooperation can be fostered - and successfully cooperative subcommunities are more likely to succeed, and have their practices spread, than noncooperative ones. Because these small communities are partially insulated from each other by *weak* network ties, they can function as quasi-independent cooperative 'innovation commons'. But because they are *connected* by weak ties, the effectiveness of cooperative internal behaviour can be observed by neighbouring communities, and thereby spread.

The model in this chapter therefore links the concept of Polanyi's republic of science with Potts' concept of the innovation commons in a way that demonstrates the potential effectiveness of a Polanyian network in fostering and disseminating Ostromian cooperation. In the next chapter, I will turn to the *epistemic* dimensions of this network, and look at the problem of modelling epistemic outcomes within the republic of science.

Chapter 6

The opinion dynamics of Mertonian science

6.1 Introduction

In chapter four I discussed the way in which the action arena of a scientific innovation commons can be located within the larger network of the 'republic of science', and in chapter five I proposed a evolutionary game-theoretic model of the cooperative behaviour of scientists within this network. However, I postponed discussion of the *epistemic* dimensions of scientific practice. It is to this issue that I now turn.

My strategy, in this chapter and chapter eight, is to begin with the simple epistemic model of scientific consensus formation proposed by Paul David, one of the founders of the 'new economics of science', and to slowly add additional components to this model until it more closely represents the scientific opinion dynamics with which this thesis is concerned. I therefore begin in section 6.2 by discussing in more detail Paul David's (1998) model of the cognitive performance of 'invisible colleges'. In section 6.3, I discuss the relation of this basic model to the framework of veritistic social epistemology discussed in chapter three, and I modify David's model to render it compatible with these metatheoretical commitments.

In section 6.4, I look at the simple dynamics of this model, both with and without observation of nature, and thus the model's representation of the movement towards consensus within scientific subcommunities. In section 6.5, I then use the model to examine two categories of problem: the relation-

ship between the scientific community and the broader public sphere, and the relationship between research sub-communities and the broader scientific community. In section 6.6, I discuss some of the shortcomings of this simple David model, in particular its inability to incorporate the epistemically important concept of priors. In section 6.7, I outline A.C.R. Martins' 'continuous opinions, discrete actions' (CODA) model of scientific opinion dynamics, which addresses these shortcomings. Finally, in section 6.8, I discuss the behaviour of this model. This lays the groundwork for the discussion of the opinion dynamics of open inputs science, in Part III.

6.2 David's invisible college voter model

David proposes his simple model of scientific opinion dynamics in his 1998 paper 'Communication norms and the collective cognitive performance of "invisible colleges" '. There David argues that 'invisible colleges' - a category that, in chapters four and five, I argued can often be understood in terms of 'scientific innovation commons' - can usefully be modelled as an unstructured graph. Each scientist or scientific research unit is linked to a series of neighbours with whom they can communicate. In David's model, each such unit is linked to three neighbours, such that the crux and three extremities of the letter Y can represent scientific research units, in a graph David represents as in Figure 1:



Figure 6.1: David's representation of the simple invisible college graph.

In this chapter, for the sake of easy compatibility with the NetLogo agent-based modelling software package, I assume rather that each scientist is linked to four neighbours, in a simple grid, as in Figure 6.2. This shift in graph structure does not materially influence the behaviour of the system. Like David, I will assume that this structure is a torus, such that every scientific agent is linked to four others.

For David, each agent can hold one of two opinions, understood as the decision to accept or reject a specific scientific hypothesis. Thus David's

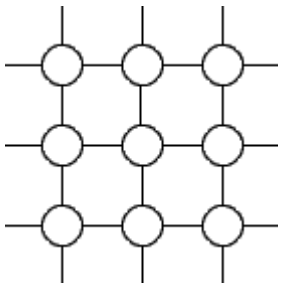


Figure 6.2: The basic invisible college network used in this chapter.

model represents a specific scientific subcommunity's debate over whether to accept or reject a given scientific claim P.

David proposes two possible categories of motivational structure for members of the scientific subcommunity. On the one hand, we can assume that scientists are motivated by truth-seeking, and rationally update their beliefs based on the best available evidence. On the other hand, we can assume that scientists are motivated by reputation, and seek to maximise their reputation within their community over the long run, by adopting beliefs that are most likely to result in scientific prestige. In line with his earlier work on the reputational incentive structure of scientific decision-making, David adopts the latter approach to modelling scientists' behaviour: for David, scientists are motivated first and foremost by reputation. In section 6.7, and then in chapter eight, I will revise this assumption, and show that dynamics similar to those of the David model can be produced even by rationally truth-seeking Bayesian agents. For now, however, I follow David in assuming that scientists are motivated exclusively by reputational outcomes.

Given this motivational structure, David considers a range of possible scenarios. He proposes the following symbolism:

> **c** denotes conformity with the preponderance of scientific opinion in one's network, and **d** denotes disagreement with that consensus. (David 1998: 138)

Following David, I denote the idea that the prevailing consensus considers a proposition to be true (Reliable/Right) as R. Similarly, I denote the idea that the prevailing consensus considers a proposition to be false (unreliable/Wrong) as W. Finally, I denote the *eventual* consensus that will emerge in the network as **R** (Right) or **W** (Wrong). A scientist considering their

reputational objectives must, in David's model, consider both the current consensus in their community, and the long-term consensus by which the current consensus will ultimately be judged. If we assume, for the sake of argument, that researchers treat the currently unknown ultimate long-term consensus as representing the true state of the world, we can schematise the possible outcomes a researcher must consider as follows (David 1998b: 139):

$\{\mathbf{c}, \mathrm{R}|\mathbf{R}\} = \mathbf{b_1}$: being right, with the crowd;
$\{\mathbf{c}, \mathrm{R}|\mathbf{W}\} = \mathbf{b_2}$: being wrong, with the crowd;
$\{\mathbf{d}, \mathrm{R}|\mathbf{R}\} = \mathbf{b_3}$: being in a minority and wrong;
$\{\mathbf{d}, \mathrm{R}|\mathbf{W}\} = \mathbf{b_4}$: being in a minority and right.

David suggests that, given these possible outcomes, scientists are likely to have predictable preferences. Clearly scientists would ultimately prefer to be right than wrong. But, as David puts it,

> it is also plausible that being wrong in a crowd will be deemed not as bad (for her subsequent reputational standing) as being found to have been wrong more-or-less on one's own; and being "lonely yet right" is deemed reputationally more glorious than being correct among a crowd of one's peers. (David 1998b: 139)

Thus David suggests that the preference ranking by scientists of these outcomes is likely to be:

$$\mathbf{b_4} > \mathbf{b_1} > \mathbf{b_2} > \mathbf{b_3}$$

Given this ranking, David then asks how scientists assess the probability of ultimate consensus forming in favour of $\mathbf{R}$.

Let Θ denote the scientist's subjective estimate of the probability of this outcome. How is Θ determined? David suggests that local consensus provides the best mechanism scientists have for estimating ultimate global consensus. From this is follows that, provided the condition $[(\mathbf{b_1} - \mathbf{b_3})/(\mathbf{b_4} - \mathbf{b_2})] > [(\mathbf{1} - \mathbf{\Theta})/\mathbf{\Theta}]$ is met, the rational strategy for any given scientist to adopt, to maximise their reputational payoff, is the strategy of conformity $\mathbf{c}$. Given the plausibility of this assumption, David concludes that, for the purposes of his model, scientists' behaviour is 'conformist'.

To maximise conformity with their neighbours' opinions, David argues, a rational actor will adopt a probabilistic strategy, selecting their own opinion

"by a procedure equivalent to tossing a coin whose loading mirrors the division of opinion among the set of her neighbors" (David 1998b: 140). This procedure is in turn equivalent to the procedure of randomly selecting one neighbour, and adopting that neighbour's opinion.

In other words, even given scientists' consideration of both short- and long-term reputational benefits, David's model will, given reasonable assumptions, reduce to the classic 'voter model' (Clifford and Sudbury 1973; Holley and Liggett 1975). That is: in any given round of play, an agent will probabilistically pick one of its neighbours with equally weighted chance, and adopt that neighbour's opinion. David's scientific invisible college model precisely maps onto this classic model.

Given this thesis's grounding in the tradition of the new economics of science, I suggest that it is reasonable to take David's model as the starting point for my own analyses of scientific opinion dynamics. In the next section, however, I will consider how well this model meets the metatheoretical commitments to veritistic social epistemology laid out in chapter three.

6.3 A veritistic approach to modelling invisible colleges

As David acknowledges, this reinterpretation of the voter model to represent the opinion dynamics of a scientific 'invisible college' is likely to trouble some theorists in its vision of the scientific enterprise. As David puts it,

> Can it be thought reasonable to so completely disregard the claims of scientists to possess other, more reliable methods, such as direct observation and experimentation, which they deploy to arrive at "closure" in discussions about the validity of scientific statements? (David 1998b: 141)

To answer this question, David reviews the sociological and philosophical debates which I briefly discussed in chapters three and four: debates within the 'strong programme' in science studies over realism and relativism, objectivity and social constructivism. As David notes, one 'strong' variant of the strong programme in science studies - what I refer to in Law and Pepperell (2017) as the 'methodological anti-realist' approach - proposes to reject the role of experimental observation altogether in analyses of the determination

of scientific belief. Against this, David contrasts a strong form of scientific realism, which rejects the role of social factors in the determination of scientific content altogether.

For David, there are positions worth defending on both 'sides' of these 'science wars'. On the one hand, it is important to analyse the social dimensions of the production of knowledge; on the other hand, it is important to acknowledge the core of the anti-relativist position. As David puts it:

> It is an important quality of "scientific inquiry" as a collective activity that it is able to escape *eventually*, even if not instantaneously, from the grip of consensus upon propositions that are falsified repeatedly by observational experience (David 1998b: 143)

This reasoning is compatible with the theoretical perspective I adopted in Part I. There I proposed a 'weak realist' epistemological approach, understood as a veritistic approach to evolutionary social epistemology.

However, there is a tension between this metatheoretical position and David's model itself. The classic 'voter model' is a model of a *closed* social system, in which agents update their opinions only in light of other agents' opinions, without observation of nature. If there is no mechanism in the model through which scientists can observe nature, then it is impossible for there to be a veritistic selection process through which an ultimately truth-oriented evolutionary process might operate. Thus, although David is committed at a metatheoretical level to a veritistic social epistemology, his model of invisible colleges cannot 'cash out' this commitment in practice.

In order to render David's model compatible with veritistic social epistemology, we must therefore incorporate observation of nature into the model.

In this chapter I do this in a very simple way. Rather than simply selecting a neighbour at random, and updating their opinion to reflect that neighbour's opinion, each agent in this chapter's model also has a certain probability α of observing nature, and a probability $1 - \alpha$ of selecting a neighbour's opinion via the process described above. Scientists who observe nature in any given round of play automatically update their opinion to reflect the content of that observation. Moreover, observations of nature are not perfectly accurate, but rather are correct with a probability of β and incorrect with a probability $1 - \beta$. These two parameters therefore determine the role that observation of nature plays in the formation of scientific opinion.

In the limit case in which $\alpha = 1$ and $\beta = 1$, all agents immediately observe nature with complete accuracy, and consensus in favour of truth

is immediately formed. In the opposing limit case in which $\alpha = 0$ (and in which the value of β is therefore irrelevant to the behaviour of the model), no observations of nature take place, and the model reverts to that analysed by David. In the middle ground between these two extremes, however, agents' opinions are determined partly by their social environment, and partly by their observations of nature. This is the modelling space in which the social opinion dynamics of a scientific subcommunity are relevant to the outcomes of an in-principle truth-oriented collective scientific enterprise.

6.4 The simple dynamics of the invisible college model

In this section I analyse the behaviour of this model using the agent-based modelling language NetLogo.[1] In the version of the model analysed in this section, I created a 20 by 20 torus of 400 agents, illustrated in Figure 3. Each agent is connected to four neighbours, by equally weighting links. The distribution of initial opinions is random, with each agent having a 50% chance of accepting or rejecting a proposition P. The update rule is as outlined in the previous two sections. In this initial version of the model, $\alpha = 0$ - that is to say, no observation of nature takes place. I ran 1000 simulations in this model, of 1000 rounds each; in each round, each agent updated its opinion once, in a randomly selected order.

This is the system and behaviour analysed by David, using the analytic tools of graph theory. As David summarises,

> It is evident on even the briefest consideration that the extremal states... in which there is a perfect correlation of beliefs throughout the population, constitute absorbing states for this system. Once such a state is entered, there can be no further change. (David 1998: 148)

This is evident in my simulations. Figure 6.4 illustrates the distribution of agents' opinions after 1000 rounds of play across 1000 simulations. Although

[1]The code used in this section and the next is available via the NetLogo modelling commons at: `http://modelingcommons.org/browse/one_model/6114`. The simulations data used to generate the graphs in this chapter is available for download from: `https://tinyurl.com/ydjl783e`.

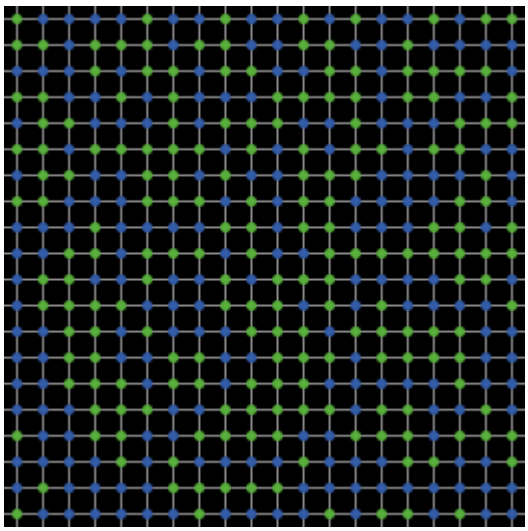


Figure 6.3: An example of the basic invisible college model in NetLogo.

are were some simulations - 37 in total - in which neither of the two extremal states has been reached, in the overwhelming majority of cases - 96.3% - the system has reached an equilibrium state, after 1000 rounds of play, in which complete consensus has been attained: either all scientists in the invisible college accept the truth of P, or all scientists reject it. In 471 simulations - 47.1% - a consensus formed around agents accepting P. In 492 simulations - 49.2% - a consensus formed of agents rejecting P. The similarity of these figures is to be expected, given that each agent begins each simulation with an equal probability of accepting or rejecting P.

This is the basic David model, as discussed in section 6.2. I now introduce observation of nature into the model, as discussed in section 6.3. The model otherwise remains the same, but with the parameters $\alpha = 0.05, \beta = 0.9$. That is to say, in any given round of play, each agent has a 5% chance of observing nature; if an agent observes nature, they have a 90% chance of observing it accurately, and a 10% chance of observing it inaccurately. Agents automatically update their opinion to reflect their observation of nature, in a round in which such an observation takes place.

Figure 6.5 shows the results of 1000 simulations run with these parameters. These results are very different from the behaviour of the model without observation of nature: outcomes strongly cluster towards belief in P. No simulations result in a majority of agents rejecting P: the simulation with

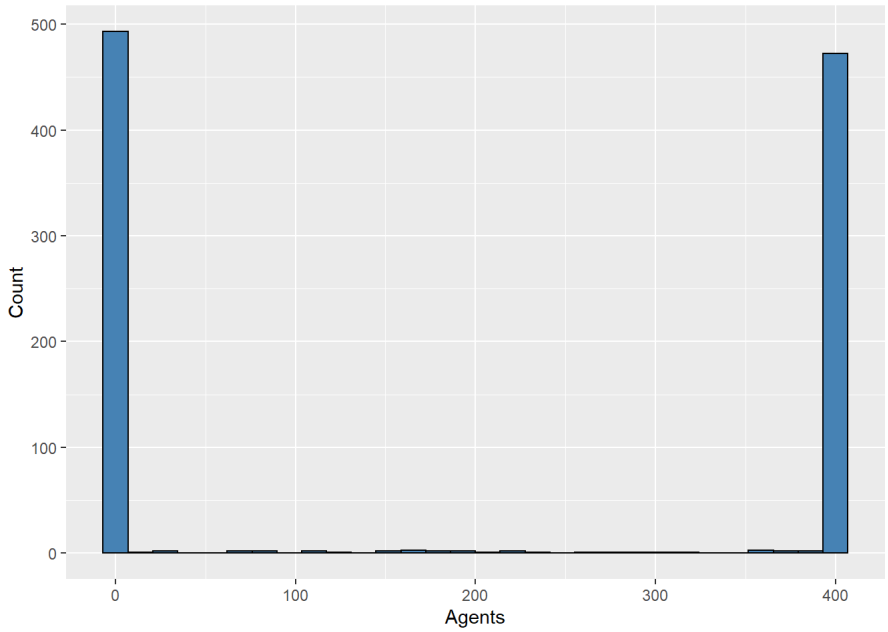


Figure 6.4: Number of agents believing P after 1000 rounds across 1000 simulations: the simple voter model.

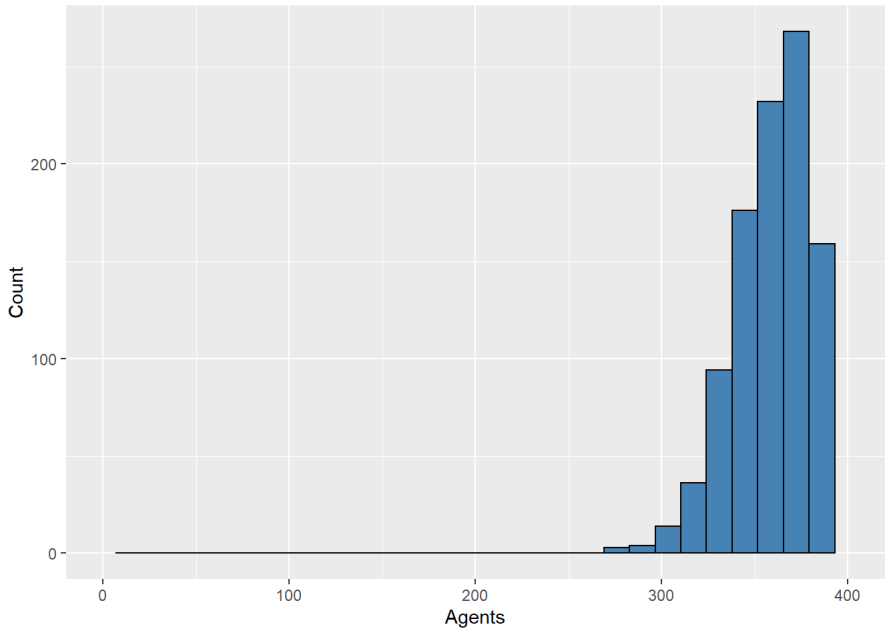


Figure 6.5: Number of agents believing P after 1000 rounds of play across 1000 simulations: with observation of nature.

the outcome furthest from consensus still resulted in 69.75% of agents (279 agents) accepting P. The median outcome, for these rounds of simulations, was 90.5% of agents (362 agents) believing P. In general, then, this level of observation of nature reliably results in a large majority of agents accepting the truth of P after 1000 rounds of play.

Unlike the previous voter model, then, the invisible college model with observation of nature does not predictably result in unified consensus. On the contrary, a significant minority of agents persist in falsely rejecting P even after 1000 rounds of play. This is to be expected, given the process via which nature is observed in the model. In the absence of wholly reliable observations of nature, the system experiences a steady input of *incorrect* information. This steady stream of faulty observations prevents the system from achieving an extremal state of total consensus. Provided, however, that β is sufficiently low, the system will result in overall outcomes which are, on average, epistemically preferable to the random outcomes of the basic voter model. Any individual scientist may be wrong, but the community as a whole is likely to be right. As might be expected, given the fallibilist nature of the scientific enterprise, a *lack* of overall consensus correlates with superior aggregate outcomes.

In summary, results from the basic invisible college model are straightforward. In the absence of observation of nature, agents will eventually unify around a consensus belief, but that belief will be determined randomly. If agents do observe nature, total consensus is much less likely, but overall opinion of the community as a whole is a reliable guide to the true state of the observed natural world.

6.5 The interaction of the college with the broader public sphere

In the previous section I outlined the simple invisible college model. In this section I start to introduce some more complexity by the model, by considering the relation between the invisible college and the broader public sphere.

The basic scenario I am considering, in this section and the next, is the relationship between a relatively small-scale invisible college of researchers engaged in active observation of nature, and a broader community of indi-

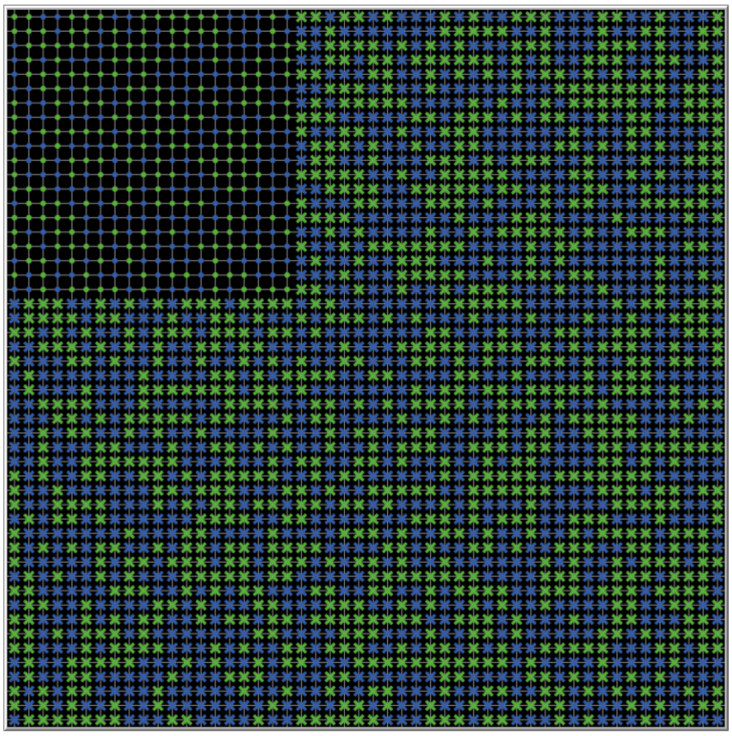


Figure 6.6: Scientists and citizens sharing a broader public sphere.

viduals who are not engaged in observation of nature, at least in relation to the relevant proposition P. In terms of the model described above, in other words, I am considering a community of agents for whom $\alpha > 0$, embedded within a community of agents for whom $\alpha = 0$. The former category of agents I refer to as 'scientists', the latter as 'citizens'. Specifically, in this section and in the next, I am considering a model, represented in figure 6.6, of 2500 agents, 400 of whom are scientists, 2100 of whom are citizens.

In the model considered in this section, scientists' update rules are the same as those described in the observation of nature model of section 4: $\alpha = 0.05, \beta = 0.9$. That is, each scientist has a 5% chance of observing nature in any given round of play. If the scientist observes nature, they have a 90% chance of observing it correctly, and a 10% chance of observing it incorrectly. If they observe nature, they update their opinion based on that observation. Citizens, by contrast, simply update their opinions in the manner of the simple David voter model.

This basic model can serve a double function in the analysis that follows. On the one hand, the model can be interpreted as a representation of the relationship between scientists and non-scientists, within the public sphere. On the other hand, the model can be interpreted as a representation of the relationship between specialists and non-specialists, within the scientific community.

As I discussed in chapter four, one of the key features of the 'republic of science' is the cognitive division of labour associated with scientific knowledge production. The differentiation of a range of scientific disciplines and sub-disciplines, and the clear demarcation of distinct research projects and paradigms within those sub-disciplines, has resulted in a high degree of specialisation, comparable in some respects to the specialisation in broader economic production analysed by Adam Smith (1976).

When we consider the relationship between a specific scientific community, and the broader public sphere, then, we should not understand this simply as a relationship between scientists and non-scientists. Rather, we are studying the relationship between those who are actively researching a specific topic, and those who are not. In the models presented here - all of which consider the opinion dynamics associated with accepting or rejecting a single proposition P - some agents are engaged in categories of observation of nature which are relevant to the evaluation of the truth or falsity of P. Other agents may be scientists engaged in other research projects, in other sub-disciplines or in other fields, who make no direct observations relevant to assessing the truth or falsity of P, but who are still part of the broader scientific community of practice. The model discussed in this section applies to this scenario, as well as to the analysis of the non-scientific public sphere.

With this dual interpretation of the model in mind, I will now consider the behaviour of the model as specified above. I ran a series of 1000 simulations in this model. Figure 6.6 represents one example of this model, at the beginning of a simulation. The smaller square of agents at the top left of the image represents an invisible college of scientists carrying out observations of P. The remaining larger group of agents represent 'citizens'. Because the graph is a torus, the location of the group of scientists in the top left of the overall community is immaterial to the behaviour of the model.

Figures 6.7 and 6.8 represent the results from these series of simulations. Comparing these two graphs shows some interesting results.

First, both communities - scientists and citizens - reliably achieve overall positive epistemic outcomes. Even though citizens cannot themselves observe

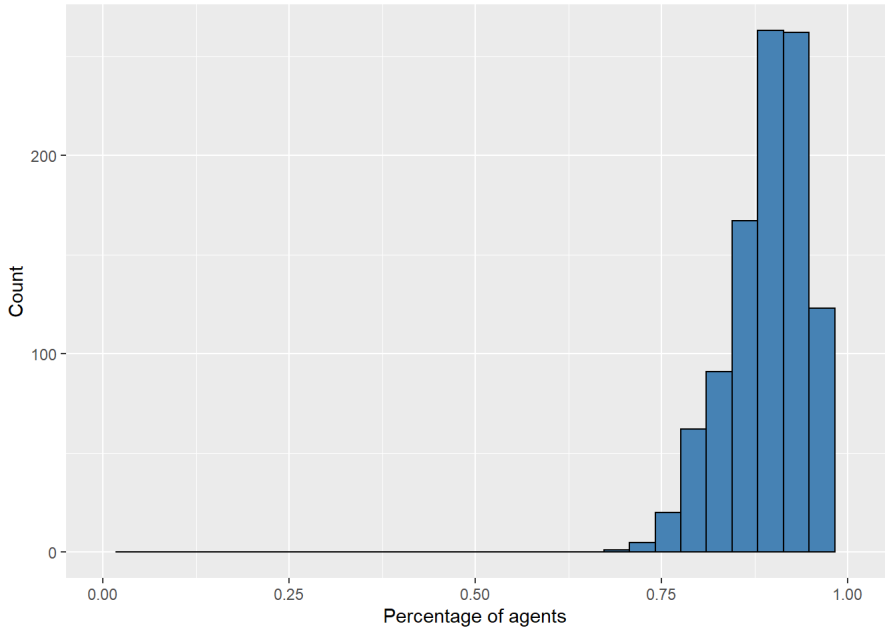


Figure 6.7: Percentage of agents believing P after 1000 rounds, across 1000 simulations: scientists within the public sphere

nature, the fact that others - scientists - in the broader community can do so, means that typically the aggregate beliefs of the community as a whole are epistemically reliable.

The scientific community is - as would be expected - extremely unlikely, in this model, to adopt overall false beliefs. None of the 1000 simulations I ran resulted in a majority of scientists rejecting the true proposition P. By contrast, 2.5% of simulations resulted in a majority of the population of citizens rejecting the true proposition P. If the majority opinion of a community is therefore treated as collective community judgement, then the community of scientists is more epistemically reliable than the community of citizens.

This conclusion is, however, strikingly dependent on our criterion of collective community judgement. This point can be illustrated by adopting a more stringent criterion of consensus: that a community has reached epistemically positive consensus if 95% or more of its members believe the truth. By this criterion, 25.6% of simulations resulted in consensus among citizens in favour of truth, but only 12.8% of simulations resulted in consensus among scientists in favour of truth.

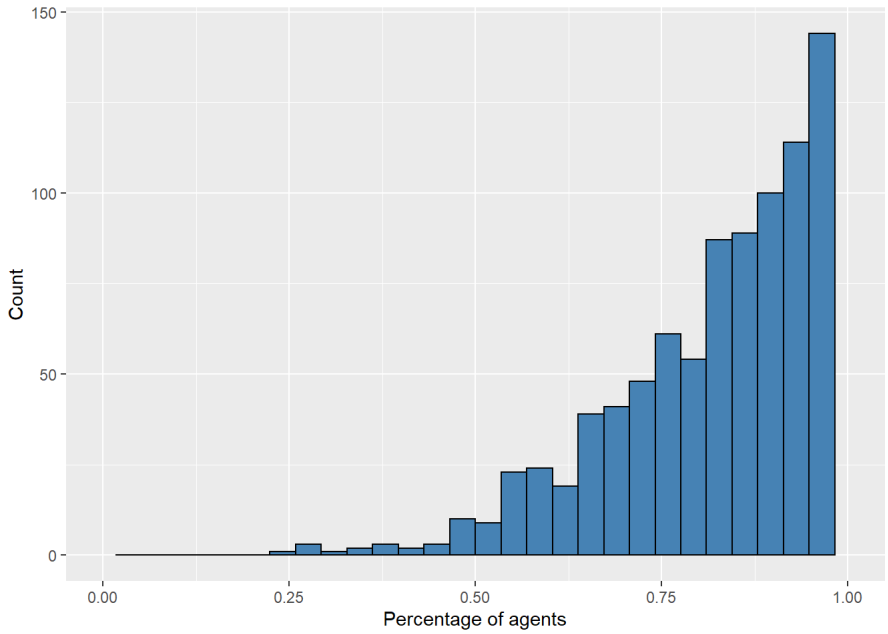


Figure 6.8: Percentage of agents believing P after 1000 rounds, across 1000 simulations: citizens within the public sphere

In other words, depending on how we assess the collective judgement of the overall community, we can potentially reach the counter-intuitive result that citizens who do *not* observe nature more reliably achieve an overall epistemic judgement in favour of the truth than scientists who *do* observe nature, *provided* those citizens exist in a network shared with a community of scientists.

This behaviour of the system is explained by the dynamics of scientist and citizen behaviour that I have discussed above. First, as noted in the previous section, the community of scientists is subject to a steady stream of false observations, while citizens are not. We would therefore expect to see a greater likelihood of very high consensus outcomes among citizens than among scientists, for the reasons discussed in section 6.4. At the same time, because citizens in this model are part of a broader public sphere also inhabited by scientists, citizens' opinions are not equally likely to move towards true or false consensus, as in the basic David model. On the contrary, the sphere of scientists, which reliably generates a distribution of beliefs heavily weighted towards truth, influences the sphere of citizens which can, in turn, potentially form a distribution of opinions even more heavily weighted towards truth, undisrupted as it is by a stream of unreliable observations.

These results are, of course, dependent on a range of specific features of the model. Different values of α and β produce different dynamics. Moreover, the analysis in this section is based on a very simple update rule for citizens. In the following section I therefore look at some of the limitations of this model. In the final section of the chapter, I will then address these shortcomings by the introduction of the Martins' 'CODA' model, which will serve as the foundation for the analysis carried out in Part III.

6.6 Limitations of the basic invisible college model

In previous sections of this chapter I have looked at the epistemic outcomes of scientific innovation commons using the tools of simple opinion dynamics modelling, building on David's 1998 model of invisible colleges. However, as I suggested in section 6.3, there are significant limitations of the David model.

A range of objections to David's basic model can potentially be articulated. These include: that the model is overly simplistic in its update rule; that it is overly cynical in its assumption that researchers are motivated

exclusively by reputational outcomes; and that it is insufficiently epistemologically realist, in its lack of role for the observation of nature. In addition, critics of the basic David model could add that treatment of the scientific community as a flat, unstructured graph does not realistically capture the complexity of actual scientific networks.

In this chapter I have addressed the objection from epistemological realism. However, the remaining objections stand. Not all of these objections can be addressed within the scope of this thesis - in particular, I do not intend to address the problem of graph structure in this thesis. The question of the *update rule* guiding agents' decisions is, however, an issue to which I will return in the next section.

One of the features of the model described in this chapter is that scientific invisible colleges do not achieve equilibrium or consensus; on the contrary, a steady stream of inaccurate observations of nature renders the scientific invisible college continuously in flux, even after an indefinite number of rounds of play. In one sense, as discussed above, this feature of the model usefully captures the fallibilism of scientific research. At the same time, however, it also suggests a limitation of the model. Although science is fallibilistic, and any scientific proposition is thus in principle open to refutation by new evidence or new reasoning, in practice scientific propositions can and often do achieve effective consensus among working scientists. Once a sufficient body of evidence has accrued, some scientific propositions are taken out of 'circulation' in the opinion dynamics of 'normal science'. Such propositions have achieved a sufficient level of community acceptance that they can become part of the accepted background knowledge for future research, rather than an active object of research themselves.

The basic invisible college model cannot capture this dynamic, and one reason for this limitation is the model's inability to represent the accumulation of evidence over time. The basic voter model, and its variants in this chapter, satisfy the Markov property of not retaining 'memory' of previous rounds of play. This feature of the model therefore intrinsically cannot capture the accretion of scientific 'memory', as evidence for a proposition accumulates over time.

Moreover, because agents have no 'memory' of previous judgements and observations, it is impossible for agents to rationally evaluate the strength and weakness of the evidence for any proposition P. This lack of memory is therefore connected to the model's 'cynicism' about agents' motivations: agents who lack 'memory' cannot easily engage in rational truth-oriented

evidence-based judgements, because they lack the background knowledge necessary to comparatively analyse and evaluate new information.

The remaining shortcomings discussed above - with the exception of the problem of graph structure - can therefore be addressed by incorporating 'memory' into agents' update rules. Modifying the David model in this way enables agents to make rational, Bayesian decisions about whether to accept or reject a proposition P based on the totality of information available to them, from both the past and the present.

In the remainder of this chapter, and in Part III below, I will aim to realise this goal, by introducing A.C.R. Martins' 'continuous opinions, discrete action' (CODA) model of scientific opinion dynamics.

6.7 The CODA model

Martins' CODA model, which will provide the foundation for my opinion dynamics modelling in the remainder of this thesis, was developed in his 2008 'Continuous opinions and discrete actions in opinion dynamics models' (cf. Martins 2009; Martins, Pereira, and Vicente 2009; Martins, Pineda, and Toral 2010), and applied to the problem of scientific consensus formation in his 2010 'Modelling scientific agents for a better science' (Martins 2010). The CODA approach can take a range of different forms. In the models discussed in this thesis, agents inhabit a regular square lattice. As in David's voter model, each agent holds one of two opinions: P or *not-P*. These values represent the agent's decision to accept or reject a specific scientific hypothesis. The models discussed in the remainder of this chapter therefore represent the dynamics of a single scientific opinion, as that hypothesis is tested and debated within a scientific subcommunity.

As discussed above, one of the shortcomings of David's voter model of scientific opinion dynamics, is that scientists' epistemic decision-making may not best be represented simply in terms of a binary choice about whether to accept or reject a given scientific hypothesis. As Martins notes -

> When someone faces a binary decision, the opinion about which option is the best one is not necessarily binary. (Martins 2008: 2)

Specifically, Martins argues that it is reasonable to see scientific agents as weighing the probability that P is true, given the totality of information

available to them, before they make the binary decision about whether to, provisionally, accept or reject P. We can therefore assume that each agent holds an individual opinion p, representing the subjective probability that P is true, and denoted by a continuous variable. We assume there is some probability value q above which agents will accept P, and below which they reject it. These two variables correspond to the *conclusions* scientists reach, and publicly express in published work, and the *priors* that inform their reasoning.

In the CODA model, then, each agent is associated with a 'private' continuous variable expressing a probabilistic judgement about the truth that P, and a 'public' discrete variable expressing their decision to accept or reject P. The model assumes that agents can observe other agents' *conclusions*, but cannot observe other agents' *priors*. Thus, agents can only update their priors in light of other agents' conclusions, or in light of their own observations of nature.

This update process is modelled in Bayesian terms. Following Martins, I represent an agent i's probability estimate that P, with the continuous variable p_i. The higher the value of this variable, the greater the agent's subjective probability estimate that P. I take 0 to be the 'tipping point' value at which an agent shifts from rejecting to accepting P, or vice versa. Thus, an agent i accepts P if $p_i > 0$. Likewise, an agent i rejects P if $p_i <= 0$.

As Martins shows (2008, 2010: 524), a Bayesian framework, on reasonable assumptions, results in a very simple social learning rule. In a given round of play, an agent i may observe a neighbouring agent j's publicly stated belief. If agents' probability estimates are represented as log odds, agents' individual update rules can be represented as:

$$p_i(t+1) = p_i(t) + sign(p_j) \tag{6.1}$$

That is to say, an agent's private probability estimates can be represented as arbitrarily large steps away from a central tipping point value at which the agent's publicly endorsed position changes.

Thus if I, a scientist, observe another scientist with opinion P, I update my internal probability estimate one 'step' in the direction of taking P to be correct. If I observe a scientist with opinion *not-P*, I update my internal probability estimate one 'step' in the direction of taking P to be incorrect. My own publicly expressed opinion depends on whether my private probability estimate of P is greater than the threshold probability above which I

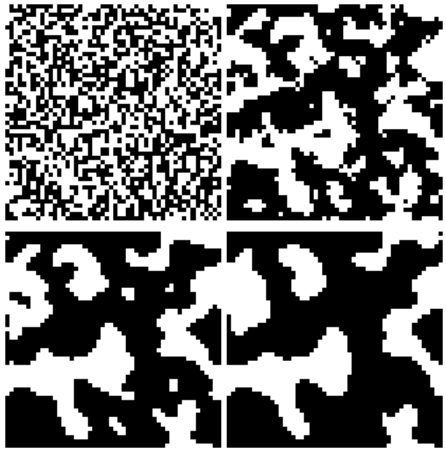


Figure 6.9: Evolution of agents' opinions

am willing to endorse the position.

The CODA model therefore reproduces the basic features of David's model of scientific consensus-formation, while adding the role of priors in scientific decision-making.

6.8 The opinion dynamics of the simple CODA model

The simplest CODA model involves agents updating their opinions based only on the opinions of their neighbours, without observation of nature. For this version of the model, in every round of play, each agent picks one neighbour at random, and updates their own opinion in response to their neighbour's opinion via the procedure described above. In the simulations described below, agents exist in a 50 by 50 torus, with unweighted links to each of their eight neighbours.[2] Each agent also begins the simulation with a private probability estimate either one step below, or one step above, the 'tipping point' value.

Figure 6.9 shows an illustrative simulation with a random initial distribution of beliefs.

[2]The code used in this section as well as in section 8.3 is available via the NetLogo modeling commons at: `http://modelingcommons.org/browse/one_model/6115`.

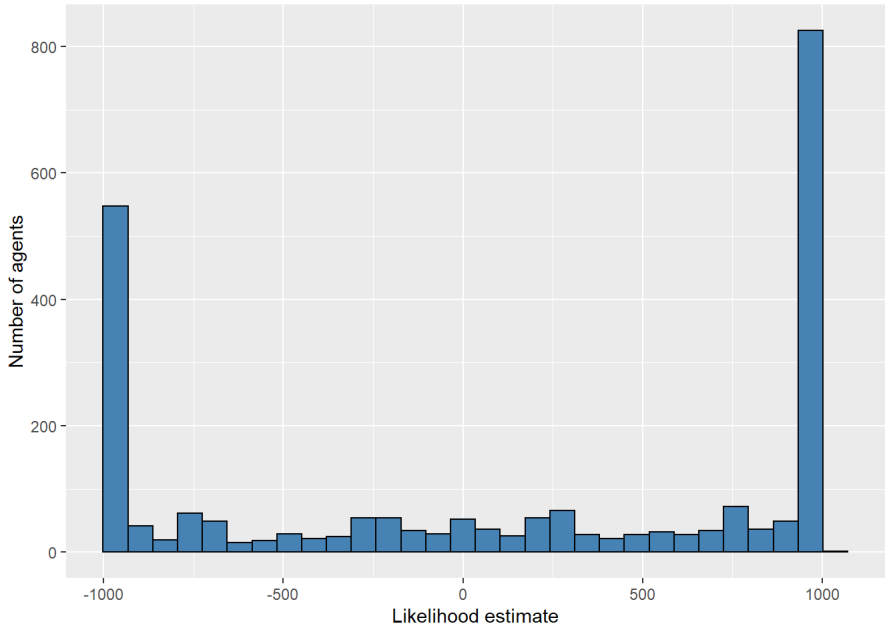


Figure 6.10: Basic CODA model: distribution of agents' likelihood estimates of P, after 1000 rounds of play

As is clear, this model quickly leads to relatively stable subcommunities of belief. Rather than a movement towards overall community consensus, the CODA model moves towards the emergence of distinct subcommunities with differing beliefs. Moreover, agents inside these subcommunities will tend to move towards 'extremist' positions. Because most agents in a subcommunity share the same belief as their neighbours, most agents within subcommunities will tend to move towards more strongly held positions over time.

Figure 6.10 represents the distribution of agents' subjective probability estimates that P, after 1000 rounds of play. As can be seen in this histogram, while there are a range of different priors held by different agents, the majority of agents hold priors at one of the two 'extremes', either strongly committed to the proposition P, or strongly committed to the proposition *not-P*.

Clearly, if we understand the goal of science as movement towards consensus, this basic CODA model represents a pessimistic vision of scientific practice. However, in this simple model agents update their opinions only in light of their neighbours' opinions, without observation of nature. To more

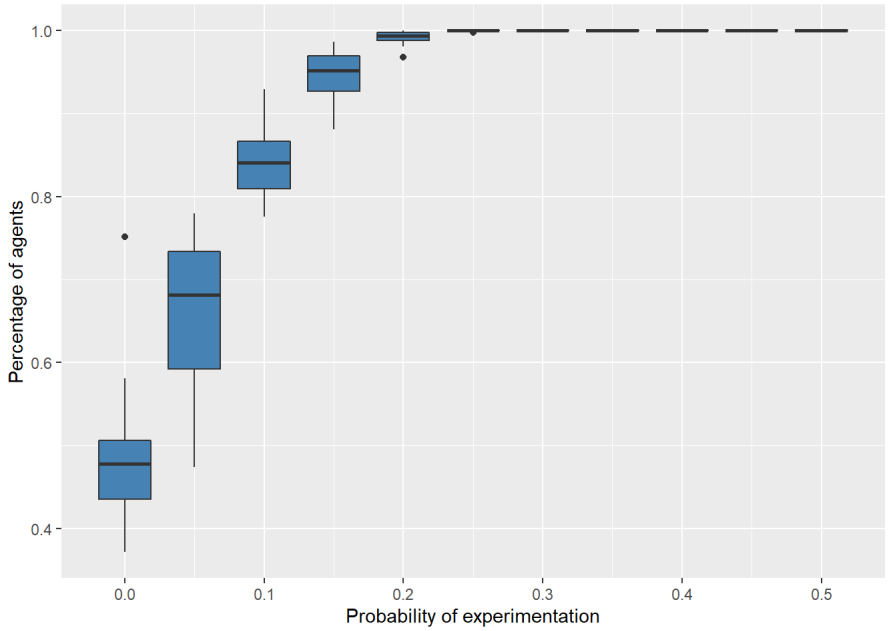


Figure 6.11: CODA model with observation of nature: Percentage of agents believing P after 1000 rounds of play, for a range of different probabilities of agents carrying out experiments

accurately represent scientific practice, we need to introduce observation of nature to the model. Just as, in section 6.3, I modified the simple David model to incorporate observation of nature, so we here modify the basic CODA model to likewise incorporate experimental findings.

Martins (2010) introduces nature by establishing a subset of agents who perfectly observe nature in any given round with probability r (Martins 2010: 524-5). In this chapter I take a slightly different approach, assuming that every agent in the subcommunity has some probability r of observing nature in any given round of play, but that such observations are imperfect, with a probability w that any given observation will produce an inaccurate report of reality. This latter assumption justifies agents' reliance on their peers in the formation of their priors.

In the simulations that follow I assume an 80% accuracy rate of observations. Figure 6.11 shows the mean proportion of agents who adopt the correct position after 1000 iterations, for a range of different probabilities of observation, given an initial equal distribution of true and false beliefs.

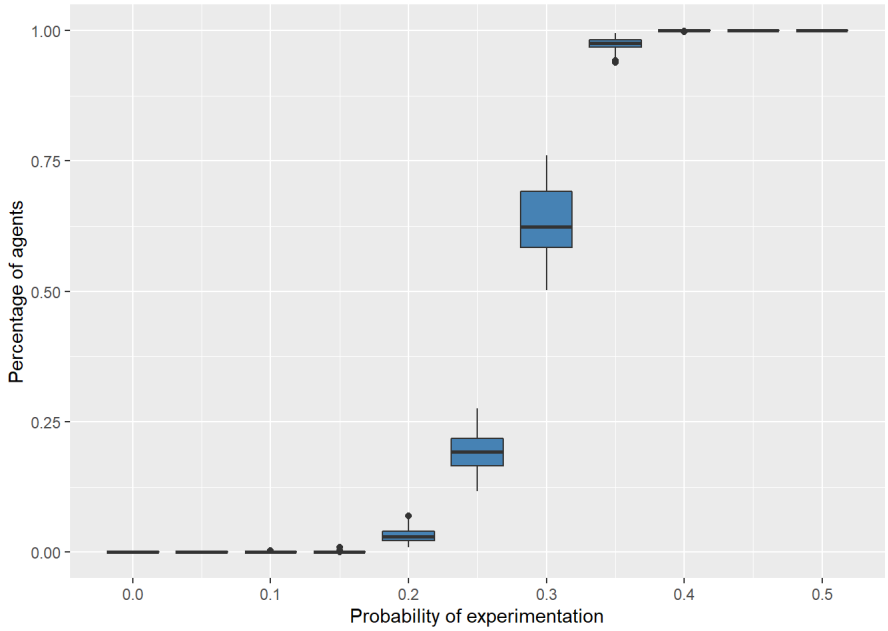


Figure 6.12: 'Pessimistic' CODA model: Percentage of agents believing P after 1000 rounds of play, for a range of different experiment probabilities, given an initial 'bad' consensus

As Figure 6.11 makes clear, even quite a low probability of agents carrying out experiments significantly increases the proportion of agents who believe the truth - greatly improving the epistemic outcomes of the system. However, quite a *high* probability of agents carrying out experiments is required for the scientific community to reach consensus after 1000 rounds of play. In my simulations, even given a starting distribution of very weakly held beliefs, every agent must have an approximately 25% chance of observing nature in each round of play, for overall consensus in favour of the truth to reliably be reached. Moreover, this scenario assumes a quite favourable starting situation, in which no agent has strong priors, and a large proportion of the community already accepts the truth.

In the next series of simulations, I model a more pessimistic scenario, in which the community again begins with weak priors, but with a consensus in favour of *false* belief. Figure 6.12 represents the results of these simulations.

In this scenario - starting from weak consensus in favour of falsehood - an experimental rate of more than 15% is required to produce even a weak

movement away from false consensus, and an experimental rate of more than 35% to produce a reliable movement towards true consensus. Again, this level of experimental observation is quite high.

As Martins notes, this 'stickiness' of scientific consensus even in response to observations that contradict it

> corresponds to the description that, quite often, old scientists do not change their minds when a new revolutionary theory appears. (527)

Instead, as Max Planck put it, scientific consensus often changes not because scientists change their minds, but because older scientists are replaced by younger ones (Planck and Rusk 1950: 33).

To represent this theory of consensus, Martins proposes a model in which scientists are gradually replaced by new scientists with moderate opinions. Again following Martins, I replicated this model by assuming that, in every round, there is a $k\%$ probability that each agent will be replaced by a fresh agent, with moderate priors. Figure 6.13 shows the results from a series of simulations in which the *experimental rate* is varied from 0 to 0.5, in increments of 0.02, and the *rate of agent replacement* varies from 0 to 0.06 in increments of 0.02.

In figure 6.13, each coloured line connects the mean percentage of agents believing P for each rate of experimentation at a given retirement rate. Thus for example the rightmost line (in red) indicates how changes in the experiment rate influence the probability of believing P, given a retirement rate of 0. The next line to the left (in green) indicates how changes in the experiment rate influence the probability of believing P, given a retirement rate of 0.02. And so on.

As this chart shows, as the retirement rate *increases*, the experiment rate required to shift agents from a 'bad' to a 'good' consensus *decreases*. With a low rate of observation of nature, social effects outweigh the observation of nature, leading to the persistence of a 'bad' equilibrium. With a high rate of observation of nature, the observation of nature outweighs social effects, resulting in collective movement towards a 'good' epistemic equilibrium. Furthermore, there is a 'step change' between these two forms of consensus. Somewhere between a 15% and a 35% rate of observing nature, the system as a whole shifts from a 'bad' to a 'good' consensus. As figure 6.13 shows, the point at which this shift happens is influenced by the rate

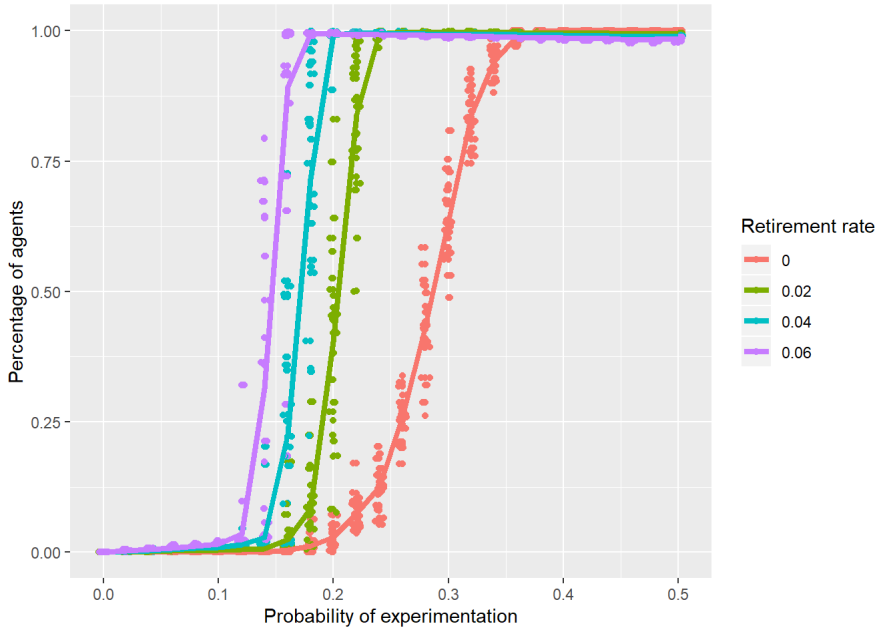


Figure 6.13: CODA model with retirement: Percentage of agents believing P after 1000 rounds of play, for a range of different experiment probabilities and retirement rates

of replacement of agents. The more rapidly agents are replaced, the lower the rate of observation of nature required to achieve an epistemically good consensus.

However, as this chart also shows, even with a quite high level of replacement of agents, a rate of observation of nature of greater than 10% is required for equilibrium to shift from a bad to a good equilibrium. Moreover, this assumes weakly held initial beliefs.

These results suggest that even in favourable conditions, including weak initial priors and a high turnover of scientists, an initial false consensus will not be supplanted by a true consensus, unless a somewhat high rate of observation of nature obtains among scientists - a somewhat pessimistic conclusion about the opinion dynamics of ordinary 'Mertonian' science.

6.9 Conclusion

In this chapter I have developed two models of the opinion dynamics of Mertonian science. These models are, in the first instance, replications of models from the literature on the economics of scientific knowledge: Paul David's 'voter model' of invisible college opinion dynamics, and A.C.R. Martins' 'CODA model' of scientific consensus formation.

I began the chapter by outlining David's simple voter model. In sections 6.4 and 6.5 I then extended David's voter model in two main ways. First, I introduced observation of nature. Second, I introduced a distinction in agent types, between 'citizens', who do not observe nature, and 'scientists', who do. The 'citizens and scientists' model enables me to represent both the relationship between a scientific and a non-scientific community, and the relationship within a scientific community between those who are engaged in research on a specific topic, and those who are not.

Studying the dynamics of the 'citizens and scientists' variant of the David model highlighted two things. First, that although a community of citizens *can* reach consensus, a community of scientists engaged in active research typically *cannot*. In this model, science is a fallibilist enterprise; part of the function of scientific research is to 'disrupt' existing dominant opinion. It is therefore unusual, in this model, for a scientific community to achieve complete consensus on a topic that is an active area of research.

Nevertheless, a scientific community engaged in active research can influence the opinions of a broader citizenry in such a way that any consensus

reached among citizens is an epistemically positive one. Scientific research communities therefore can, in this model, serve the apparently paradoxical function of enabling a citizen sub-community consensus in favour of truth, which the scientific sub-community itself does not achieve.

Having outlined these dynamics of the 'citizens and scientists' variant of the David voter model, I then highlighted some of the model's flaws. In particular, I argued, the model's agents' lack of 'memory' leaves the model unable to capture some of the most important elements of scientific decision-making. Specifically, the David model cannot capture the crucial importance of 'priors' in the scientific decision-making process.

To address this shortcoming, I introduced A.C.R. Martins' 'continuous opinions, discrete actions' (CODA) model of scientific opinion dynamics. Martins' model is very close to David's in many respects, but it introduces an additional layer of scientific decision-making: each scientist possesses a subjective probability estimate of the likelihood of the truth of any given scientific proposition. In the CODA model, scientists can therefore update their subjective probability estimate on Bayesian principles, given their observations of other scientists' findings and of their own experimental results.

The dynamics of this CODA model are very different from the dynamics of the simple David model. Scientists quickly form sub-communities of belief, and members of those sub-communities hold increasingly 'extremist' convictions, except at the communities' margins. This dynamic can be offset by observation of nature - but quite high rates of observation of nature are required, to offset social herding effects. Even with the regular replacement of 'extremist' scientists by more moderate new agents does not undo the need for quite high rates of observation of nature.

This 'pessimistic' result is where I conclude the chapter. Unlike the 'scientists and citizens' model, discussed earlier, this CODA model is not a novel contribution to the literature. However, outlining the CODA model and its dynamics sets the stage for Part III of the thesis, where I will build on this foundation to propose a novel model of 'open science' opinion dynamics.

This discussion concludes the chapter's discussion of the opinion dynamics of Mertonian science. It also concludes Part II of the thesis.

Throughout Part II, I have developed a series of models of traditional 'Mertonian' science, building on the core ideas of the new economics of science and the economics of scientific knowledge. Each chapter of Part II has

contributed a different element to a broader picture of Mertonian science.

In chapter four I outlined an Ostromian analysis of the creation of a paradigmatic scientific 'action arena': the institutional formation of the 'rules of the game' of scientific research.

In chapter five, I built on the insight of the new economics of science, that scientific input sharing could usefully be modelled as an n-player prisoner's dilemma. I extended this idea to develop an evolutionary game-theoretic model of scientific cooperation within the 'republic of science'.

Finally, in this chapter, I built on work in the economics of scientific knowledge, to elaborate two models of scientific opinion dynamics.

Taken all together, these models provide an interlocking set of pictures of different aspects of Mertonian science. These models all draw and build on existing work in the economics of science; in doing so, they also provide a complex composite picture of scientific institutional dynamics and incentive structures.

Finally, in addition to their own original contributions to the literature, these chapters also provide the background for Part III of the thesis. In Part III, I will extend the models of science developed in these chapters, to examine the impact of the institutional changes associated with 'open inputs science'.

Part III

Chapter 7

The rules of the open inputs game

7.1 Introduction

In Part II of the thesis, I developed a series of models of Mertonian science. In doing so, I built on a series of existing models: Ostromian Institutional Analysis and Development analyses of scientific knowledge commons; Dasgupta and David's analysis of science as an n-player prisoner's dilemma; and opinion dynamics models of scientific consensus-formation. Using these approaches together, I established a picture of the rules and dynamics of Mertonian science that built on, but went beyond, existing analyses in the 'new economics of science'.

My goal in Part III of the thesis is to build on the analysis of Part II, to engage in the *comparative* analysis of Mertonian and 'open inputs' science.

Part III of the thesis therefore extends the models developed in Part II, to the study of open inputs science. In doing this, I recapitulate the structure of the argument of Part II. First, in this chapter, I discuss the 'rules of the game' of open inputs science, and the ways in which those rules of the game shape the structure of the scientific action arena. Then, in chapter eight, I look at the implications of these shifts in the rules of the scientific game for the dynamics of scientists' behaviour. Just as in Part II, I use agent-based models to look at both the behavioural dynamics of scientific cooperation, in the form of input sharing between scientists, and the opinion dynamics of scientific consensus-formation. Finally, in chapter nine, I review the findings of the thesis, and draw out some of their implications for current debates around open inputs science.

7.2 Debating open inputs science

In Part I of the thesis I discussed the institutional proposals associated with open inputs science - the broad sharing in the open public sphere of intermediate inputs of scientific production. I briefly discussed some of the key examples of open inputs science: open data; open source code; open lab notebooks; and open research protocols. I also discussed the key arguments in favour of open inputs science put forward by its advocates. The two most central of these are arguments from *efficiency* and from *reliability*. In this section I will look in more depth at these arguments, and recent debates about them.

7.2.1 Open inputs and scientific efficiency

One of the key arguments for open inputs science has been the argument for efficiency. As I have explored in earlier chapters, scientific research is a collective process that operates via a complex division of labour. The sharing of scientific *outputs*, described by Polanyi (2000), allows scientists to update their own opinions in response to the findings and arguments of other scientists. This collective process sifts and evaluates diverse scientific hypotheses and evidence, thereby generating an aggregate opinion of the scientific community which transcends the opinions of any individual scientist. This traditional Mertonian scientific incentive structure however only permits this process to operate at the level of scientific *findings*; open inputs advocates suggest that a similar process can operate at the level of scientific *inputs*.

There are two key ways in which scientists sharing the inputs associated with this division of research labour may increase the efficiency of the scientific research process. First, researchers sharing the intermediate inputs of scientific production may reduce the duplication of work involved in scientific discovery. Scientists able to draw on other researchers' data, code, etc. do not need to produce that data or code themselves in order to participate in the discovery process; sharing inputs therefore increases the speed and reduces the labour of the scientific production process. As Brase puts it, the absence of open input sharing "causes unnecessary duplication of research efforts" (Brase 2014: 327).

Second, pooling resources allows scientists to exchange information and ideas in a way that may generate insights and innovations that would oth-

erwise not exist. That is to say, scientific information sharing can increase the creativity of the scientific research process, because making broader intermediate inputs available "allows research synergies" that might otherwise not be possible. (Fecher, Friesike, and Hebing 2015: 26; cf. Nielsen 2012: 85)

Thus for example, Woelfle, Oliaro and Todd (2011) have argued that their creation of an open information sharing forum for biomedical researchers serves as a "research accelerator" (Woelfle, Olliaro, and Todd 2011: 748) which increases scientific productivity. Similarly, Frankenhuis and Nettle argue that open science practices:

> can foster creativity because they cultivate an open and flexible mindset, create a more collaborative and constructive climate, and generate more accurate information and make it more accessible. (Frankenhuis and Nettle 2018: 439)

For all these reasons, open inputs advocates argue that open data sharing is key to increasing scientific productivity. Scientific projects are completed more quickly and efficiently, and forms of scientific innovation may take place that would otherwise be impossible.

7.2.2 Open inputs and scientific reliability

At the same time, many debates around open inputs science relate to the issue of scientific reliability. Here the key argument is that sharing scientific inputs allows other researchers to better evaluate the evidence and reasoning behind scientific findings: an *epistemic* argument for open inputs science.

The epistemic case for open input sharing has several elements. In the first place, open sharing of intermediate inputs reduces the opportunities for and increases the difficulty of research fraud. This benefit of open inputs has been foregrounded in recent years by several of high profile scandals - the cases of Diederik Stapel (Stroebe, Postmes, and Spears 2012), Michael LaCour (Van Noorden 2015), Woo Suk Hwang (Chong and Normile 2006), and others (Lin 2013; Gewin 2012) have highlighted the risks to the broader scientific knowledge base of scientific deception.

In Stroebe, Postmes and Spears' words, such cases have led many scholars to suggest:

> alternative mechanisms for fraud detection that need strengthening, principally greater transparency with data, including deposit-

> ing data in repositories where they can be accessed by other scientists... Inventing a whole data set is time-consuming and risky. Even though this requirement would not prevent researchers from fabricating or falsifying data, the fact that they usually make mistakes would aid detection. (Stroebe, Postmes, and Spears 2012: 682)

Beyond the issue of fraud, open input sharing also reduces the opportunity for error. As Freese and Peterson put it:

> Researchers who can keep anyone else from verifying results are insulated against any definitive demonstration of error. (Freese and Peterson 2017: 153)

Moreover, the knowledge that scientists' research inputs will be shared may incentivise a salutary care and rigour even if observation of the inputs by peers never in fact takes place. Thus Easterbrook argues that the difficulties in making code "truly open source - in the sense that it can be run, usefully, by a wide variety of people and on a wide variety of platforms" are sufficiently substantial that this is not a realistic expectation for many research projects. (Easterbrook 2014: 780) Nevertheless:

> Asking authors to open up access to their code is likely to lead to a rapid improvement in code quality: the mere possibility that someone could read the code is a strong incentive to make it presentable (even if nobody does read it in the end). (Easterbrook 2014: 781)

In recent years the debate over the epistemic benefits of open inputs science has increasingly come to focus on the issue of *replication* and *replicability*. This focus is to a large extent due to anxieties over the so-called 'replicability crisis' afflicting the legitimacy of findings in some scientific fields (Earp and Trafimow 2015; Everett and Earp 2015; Shrout and Rodgers 2018). This 'crisis' has several dimensions, but it groups together a range of worries over the reliability of the scientific literature, prompted by the failure to replicate prominent findings. In 2015, a group of researchers working as 'the Open Science Collaboration' carried out replications of 100 studies from three prominent psychology journals, and found that only 36% of replications had statistically significant results (Open Science Collaboration 2015). Along

similar lines, Camerer et al. (2016) attempted to replicate 18 experiments published in the *American Economic Review* and the *Quarterly Journal of Economics*, and concluded that only 11 (61%) of those papers were replicable, typically with a smaller effect size.

Such failures of prominent efforts to replicate important research findings have led many practitioners to worry about the overall reliability of our scientific knowledge base.

Weaknesses in the scientific production process, to which the replicability crisis has drawn scholars' attention, include fraud and uncheckable error, but also the 'file drawer problem': the risk that researchers' and publishers' preferences for 'successful' studies with positive findings result in published studies being "a biased sample of the studies that are actually carried out" (Rosenthal 1979: 638). As Rosenthal puts it, in the paper that coined the term:

> The extreme view of this problem, the "file drawer problem", is that the journals are filled with the 5% of the studies that show Type I errors, while the file drawers at the back of the lab are filled with the 95% of the studies that show nonsignificant (e.g. $p > .05$) results. (Rosenthal 1979: 638)

More recently, attention has focussed on a set of problems that Simmons et al. refer to as "researcher degrees of freedom" - the ability of researchers to make a large number of methodological decisions as they carry out their research that knowingly or unknowingly 'tilt the balance' of their findings towards the outcomes they desire (Simmons, Nelson, and Simonsohn 2011: 1359; cf. Gelman and Loken 2013). In 1998 Norbert L. Kerr coined the acronym HARKing - "hypothesizing after the results are known" - to describe the practice of retrofitting 'hypotheses' to match unexpected findings, thereby giving the false appearance of rigorous confirmatory research to findings that are in reality 'exploratory' (Kerr 1998). Even with pre-specified hypotheses, however, many "researcher degrees of freedom" can remain. In Simmons et al.'s words:

> In the course of collecting and analysing data, researchers have many decisions to make: Should more data be collected? Should some observations be excluded? Which conditions should be combined and which ones compared? Which control variables should

> be considered? Should specific measures be combined or transformed or both? (Simmons, Nelson, and Simonsohn 2011: 1359)

Because it is uncommon for researchers to make such decisions in advance, researchers have considerable leeway, when carrying out the research, to modify the research in ways that are more likely to produce statistical significant or otherwise desirable findings. This process can happen without any nefarious intent on the part of the researcher. As Simmons et al. note:

> people are self-serving in their interpretation of ambiguous information and remarkably adept at reaching justifiable conclusions that mesh with their desires. (Simmons, Nelson, and Simonsohn 2011: 1359-60)

An empirical study of psychologists carried out by John, Loewenstein and Prelec in 2012 also found a large proportion of anonymous respondents willing to admit to 'questionable research practices' (QPRs). In John et al.'s study, more than 60% of respondents admitted to "in a paper, failing to report all of a study's dependent measures", and more than 55% admitted to "[d]eciding whether to collect more data after looking to see whether the results were significant" (John, Loewenstein, and Prelec 2012: 525). As the authors of the study put it:

> QPRs are the steroids of scientific competition, artificially enhancing performance and producing a kind of arms race in which researchers who strictly play by the rules are at a competitive disadvantage. (John, Loewenstein, and Prelec 2012: 524)

Here the collective action problem of open inputs science extends beyond to the lack of collective benefits associated with input sharing, to an actively destructive outcome: bogus findings disseminating across the literature because research best practice is a strictly dominated research strategy.

Findings such as these have led to a pessimism in much of the science studies literature about the reliability of our scientific knowledge base. This pessimism has been most notably articulated by Ioannidis's argument that "most published research findings are false" (Ioannidis 2005). Ioannidis proposes a simple calculation, which uses the three variables of a posited ratio of 'true relationships' to 'no relationships' among hypotheses in a given field, the 'false negative' rate, and the 'false positive' rate, to calculate the positive

predictive value of a research finding - that is to say, the probability that any given research hypothesis will actually be true, given that it has been found to be true in a published paper. Ioannidis then shows that it is difficult to devise plausible values for these variables that do not result in the conclusion that more than half of published research findings are false.

Such arguments, and the broader debate over scientific reliability, have fostered a renewed interest in institutional *solutions* to the replicability crisis - and specifically the problems of scientific fraud, error, motivated research design, and selective publication. Open inputs science has been a common proposal from those seeking such solutions. Shrout and Rodgers are representative when they "recommend that researchers adopt open science conventions of preregistration and full disclosure" as a solution to the challenges of the replicability crisis (Shrout and Rodgers 2018; cf. Nosek et al. 2015).

I will return to these epistemic arguments in favour of open inputs science in more detail in section 8.3. In the next subsection, I will briefly discuss some of the epistemic arguments that have been made *against* open inputs science, before going on to discuss practical obstacles to these institutional proposals, in section 7.3.

7.2.3 Evaluating open inputs science

In the previous section, I reviewed a set of arguments for open inputs science. These arguments are not, however, universally accepted. Some scholars have argued against elements of open inputs science in principle. Thus for example Levy and Johns argue that open data can be a "Trojan horse" via which special interests can hope to confound regulatory policy based on scientific findings. Levy and Johns warn that:

> open-data policies might, paradoxically, *increase* industry "capture" of regulatory processes, as resource-rich special interests exploit scientific uncertainty to impose undue administrative delay. (Levy and Johns 2016:3)

From this perspective, the risk of 'motivated reasoning' in data analysis is not reduced by making data available to inspection. Rather, expert and impartial scientists are more likely to generate reliable data analysis than special interests or amateurs who might take advantage of open inputs science to muddy the waters of legitimate scientific analysis (cf. Ramírez-i-Ollé

2015). This is the same fear expressed by some critics of the broader open science movement, who worry about an erosion of scientific authority and a corresponding decline of scientific credibility as the distinction between expert and amateur is diminished (cf. Gieryn 1983, 1999).

Some scholars and practitioners have also expressed reservations about the impact that global resource inequalities may have for open science as practiced. Thus Serwadda et al. warn that:

> Open data effectively makes the resources collected in one country freely available to others. African scientists have expressed concern that open data compromises national ownership and reopens the gates for "parachute research" (i.e., Northern researchers absconding with data to their home countries). (Serwadda et al. 2018: 642)

For these practitioners, the basic incentive problem of intermediate input sharing is compounded by the inequities associated with the global distribution of scientific resources, leading to the worry that, for researchers in the Global South, "that which has taken you like 2 months, they can come and do it overnight with a whole research lab" (Bezuidenhout 2019: 22).

Within the social epistemology literature, and as discussed in chapter two, Zollman (2007) has recently emphasised the potential epistemic benefits that can counter-intuitively be associated with *limited* information sharing among working scientists (cf. Grim 2009). Zollman's findings suggest that faulty scientific consensus can easily be reached if scientists too rapidly update their own beliefs in the light of others' research, thereby reducing the cognitive diversity upon which the Popperian dynamic of epistemic evolutionary competition is based. This finding suggests that in principle open science, including open inputs science, could *reduce* the epistemic reliability of our knowledge base, by facilitating rapid consensus formation around flawed findings or data.

I will return to some of these arguments in chapters eight and nine. Most arguments against open inputs science, however, do not contest that the broad sharing of intermediate inputs has some value. Rather, critics more commonly emphasise the substantial *practical* or *institutional* obstacles to open inputs science. In the next section I therefore look at four of the most important obstacles to open inputs science, in order better to evaluate their relevance to the questions considered in this thesis.

7.3 Obstacles to open inputs science

Historically, by far the greatest obstacles to open inputs science have been practical: until relatively recently, it has simply not been technologically feasible to place many categories of scientific input into the public sphere.

In recent decades, however, the technological constraints on information sharing have been transformed by the rise of information technologies.[1] Greater data storage, processing, and dissemination technologies have made it possible for new kinds of research to be conducted, and for scientists to transfer huge amounts of information at relatively low cost, independent of both the formal process of scientific publication and the informal 'tacit' transmission mechanisms of training and collaboration. As Cameron Neylon and Shirley Wu put it:

> The Internet and the World Wide Web provide the technical ability to share a much wider range of the evidence, argument and conclusions driving modern research.... The rapid expansion and development of tools that are loosely categorised under the banner of 'Web 2.0' is what makes the sharing of research material practical. (Neylon and Wu 2009: 1-4)

In the more hyperbolic words of the Science International Group:

> The digital revolution of recent decades is a world-historical event as profound and more pervasive than the introduction of the printing press. It has created an unprecedented explosion in the capacity to acquire, store, manipulate and instantaneously transmit vast and complex data volumes. (Science International 2015: 3)

This technological shift is a condition of possibility of open inputs science: by reducing the transaction costs of input sharing, the space of feasible institutions has been greatly increased. In Djankov et al.'s term, the reduction in transaction costs has shifted the "institutional possibilities frontier" to encompass a far larger institutional space (Djankov et al. 2003).

[1] For a survey of the rise of new data technologies, see Walker (2014). For discussion of the relation of technological change to open science, see Bartling and Friesike (2014). For a discussion of the conflictual institutional possibilities of internet technologies, see Pepperell and Law (2015).

However, as I discussed in chapter three, shifts in transaction costs cannot in themselves explain institutional changes. Rather, institutional changes are better understood as changes in the 'rules of the game' that govern economic life, with transaction costs partly influencing and partly determined by those shifts. For present purposes, what is of interest is therefore the rules of the scientific game - normative, regulatory, explicit, tacit - that constrain scientists' ability or desire to engage in open inputs science *given* the technological feasibility of such participation.

Several distinct categories of institutional constraints on 'open inputs science' are relevant to this problem space:

- **Ethical constraints.** Scientists may be prohibited from sharing intermediate inputs due to ethical constraints associated with the research design. For example, research participants may have participated in the research only because the anonymity or confidentiality of the data they contributed was guaranteed by the researchers, or the data may have been shared with the original researchers under embargo for a certain period. Much medical research is carried out using datasets that cannot be shared in their entirety for ethical reasons; similarly, many categories of social research concern sensitive topics such that data cannot be shared with researchers outside the original research team. These kinds of ethical issues often place hard constraints on the degree of intermediate input sharing in which researchers can engage.

- **Coordination challenges.** There are coordination challenges associated with intermediate input sharing. Code written in a specific programming language may not be easily accessible to research teams that use a different language; data presented in a specific format may not be accessible to teams familiar with alternative data presentation. Moreover, documentation may be necessary to enable any researcher to make sense of inputs shared, which increases the costs of making materials available, imposing an extra step in the scientific production process. One of the challenges of open inputs science is resolving these coordination problems in a way that makes widespread sharing feasible.

- **The cost of input sharing.** Even with much lower transaction costs than has historically been the case, input sharing costs may remain substantial. Creating and maintaining repositories for code, data, research protocols, etc. can be costly. Moreover, data, code, lab notes,

research protocols, or any other forms of scientific intermediate input are likely to need to be modified from the forms in which they are used internally within a research team, in order to be usefully accessible to other researchers. This process of making intermediate inputs transparent and accessible may be costly or time consuming.

- **Intellectual property rights.** The basic Mertonian model of science that has informed much of this thesis assumes that scientists are exchanging their entitlement to intellectual property rights in their discoveries for 'reputational property rights' derived from priority of publication. For the scientific institutions, both historical and contemporary, which Merton studied, this assumption was often a robust one. However, in the decades that have followed the publication of Merton's major works, the norms around intellectual property rights in science have shifted, to make intellectual property rights more central to the scientific production process (Nelson 2004; Berman 2011; Mirowski 2011). These shifts in scientific norms and practice have made intellectual property rights a more important constraint on scientific intermediate input sharing than in the middle decades of the twentieth century: scientists may be reluctant to share intermediate inputs because either their institution or other organisations with which they are collaborating have intellectual property rights over the knowledge resources associated with scientific production.

Each of these constraints poses important practical challenges for implementing intermediate input sharing for scientific research teams. However, this thesis - and the remainder of part III - is focused on a different kind of impediment to intermediate input sharing: *reputational* costs and constraints. For the remainder of the thesis I therefore bracket the obstacles to intermediate input sharing associated with ethical constraints, intellectual property rights, etc., and focus almost exclusively on the problem of the *reputational* economics of open inputs science.

7.4 Reputational constraints on intermediate input sharing

In Part II, I developed a set of models of intermediate input sharing *within* small groups of scientists that I characterised as 'innovation commons'. I

argued that innovation commons are scientific research communities that can overcome the Mertonian disincentives to intermediate input sharing, by developing a set of shared norms and informal governance principles within an 'Ostromian' commons. However, even if the dynamics I analysed in Part II can incentivise input sharing *within* a scientific innovation commons, this does not intrinsically result in input sharing *beyond* the scientific innovation commons - that is, in open inputs science. The Mertonian disincentives to input sharing that I discussed in Part II in relation to scientific innovation commons are also operative when we turn now to *open* inputs science.

As open inputs science has become more prominent, more research has been carried out on the implications of Mertonian incentive structures on open input sharing. A recent article by Nelson, for example, looks at the strategies that scientists adopt to manage the tensions between different scientific norms: those of sharing and of secrecy (Nelson 2016). Nelson enumerates a range of strategies that scientists adopt to manage this tension: scientists may patent key pieces of information; they may delay disclosure of information, in order to temporarily gain the advantages of secrecy; they may choose to share inputs only with a small circle of trusted colleagues; or they may engage in "strategic withholding", whereby some information is shared, but not enough to permit the information to be mobilised by rival researchers.

The use of such strategies by working scientists is not new. In a 1974 *Science Studies* article, for example, Harry Collins studied the scientific community engaged in research into the 'TEA laser' (the Transversely Excited Atmospheric Pressure CO2 laser). As one of Collins' interview participants put it:

> What you publish in an article is always enough to show that you've done it, but never enough to enable anyone else to do it. If they can do it then they know as much as you do. (Collins 1974: 176)

The goal of sharing enough knowledge to "show that you've done it", without sharing enough knowledge to "enable anyone else to do it" can be a difficult balancing act. One of Nelson's interview participants discusses a researcher who made the mistake of sharing crucial information that enabled a rival research team to achieve priority of discovery:

> The whole race [for isolating interferon, a key protein] was won when a scientist from Genentech attended a meeting, and he copied down the first four amino acids of fibroblast interferon which were flashed on a screen by one of the scientists who was working on interferon in Japan. With that information, we were immediately jumping ahead of the competition, because we were immediately synthesizing the DNA and giving the molecular biologists the unique tools to go out and fish out the gene.... In retrospect, the presenting scientist will probably never forgive himself. But... It wasn't until later that it became apparent that a small piece of information - but crucial piece of information - can start a snowball effect if you have the technology to turn that bit of information into a unique strategy that nobody else can adopt. (Crea 2004, quoted in Nelson 2016: 279)

Such disincentive to open inputs sharing may be particularly strong in contexts of commercial competition, but they can apply in *any* research context structured by the reputational incentives of Mertonian science. For advocates of open inputs science, then, one of the key challenges that must be overcome is this Mertonian incentive structure that rewards scientific input hoarding. In the next section, I therefore look at the approaches that have been adopted by policymakers and practitioners to try to foster open inputs science in this Mertonian context.

7.5 Open inputs policies

So far I have described a range of different obstacles to open inputs science. In this section I look at some of the ways in which policymakers and practitioners have sought to overcome these obstacles, by incentivising open inputs science. Broadly speaking, these strategies can by typologised along two dimensions. In the first place, interventions take place at different scales of social action: interventions into the incentive structure of science can be made by national and transnational governance institutions; academic institutions: professional bodies; universities; academic publishers; etc.; and 'on the ground' by practitioners. Initiatives can also broadly be categorised into three streams: those that function as *facilitators* for open inputs science; those that function as *incentives* ('carrots') and those that function as *dis*incentives ('sticks').

Thus for example the European Commission's recent development of the European Open Science Cloud (EOSC) (European Commission 2017, 2018) aims in significant part to develop the infrastructure required for large scale scientific input sharing within the 'digital single market', and thereby facilitate greater research input sharing. The EOSC ultimately aspires to establish:

> a virtual environment with free at the point of use, open and seamless services for storage, management, analysis and reuse of research data, across borders and scientific disciplines. (European Commission 2018: 16)

Although the primary goal of the EOSC is "to give Europe a global lead in scientific data infrastructures" (2018: 16), the EOSC also aims to be capable of integrating into "a global federation of federated services" (2018: 19), as part of a broader shift towards more integrated and open global research infrastructure. The EOSC thereby connects to other initiatives, like the NIH Data Commons (NIH 2018), the Australian National Data Service's Research Data Commons (ARDC 2019), the African Data Intensive Research Cloud (Simmonds et al. 2016), and others (National Academies of Sciences and Medicine 2018; Open Research Data Task Force 2018).

These projects aim to permit researchers to share data and other research inputs more easily, both by providing a physical and governance infrastructure, and by establishing a set of shared standards, norms, and practices that can greatly reduce the coordination and 'translation' costs involved in sharing scientific inputs across different projects, disciplines, nations, and frameworks. One example of this is the widespread adoption of the FAIR framework for digital object management and stewardship. (Wilkinson et al. 2016; Mons et al. 2017) The FAIR acronym seeks to capture a range of desirable attributes of scientific information. For FAIR advocates, scientific information should seek to be: Findable; Accessible; Interoperable; Reusable. Implementing these principles requires the creation of a globally unique identifier for each digital object; searchable metadata linked to those identifiers; clear licencing rules for digital objects; and a range of other actions (Wilkinson et al. 2016: 4). Moreover, although these principles are most often discussed in relation to the most commonly shared form of scientific intermediate input - data - they can be applied to any knowledge input

(European Commission 2018: 8).[2]

These national and transnational infrastructure projects intersect with private initiatives like Zenodo (Zenodo 2019), Figshare (Thelwall and Kousha 2016), Dryad (White et al. 2008) and others, to provide a wide range of infrastructure resources for input sharing.

Open science policy has therefore facilitated input sharing both by the construction of infrastructure, and by contributing to the solution of the coordination problem of establishing shared standards and protocols. At the same time, policymakers have sought to incentivise open science and disincentivise closed science practices, beyond the creation of material and intellectual resources.

Policymakers have increasingly instituted *rules* for input sharing and stewardship, using the levers of public funding to mandate forms of scientific input sharing among publicly funded researchers. The European Commission thus proposes that:

> in the context of Open Science initiatives, the FAIR principles should be implemented in combination with a policy requirement that research data should be Open by default. (European Commission 2018: 21)

Similarly, the US National Institutes of Health aims to

> increase the value of NIH-supported data by democratizing access and use of data and analytical tools (National Institutes of Health 2018: 9)

At the same time, policymakers have sought to shift the *incentives* around scientific input sharing. As discussed in section 7.4 above, and throughout the thesis, the relative lack of reputational dividend associated with sharing intermediate inputs is a key problem in encouraging open inputs science. Policymakers have therefore sought to incentivise input sharing by encouraging the development of alternative research metrics that give greater reputational weighting to input sharing. Since 2013, for example, the US National Science Foundation has asked principal investigators

> to list his or her research 'products' rather than 'publications'... Data sets, software and other non-traditional research products will count too. (Piwowar 2013: 159)

[2]See Mons (2018) for an overview of implementing FAIR data principles in practice.

Policies have also been shifting in many professional and scholarly institutions. The Royal Society, often regarded as the oldest scientific institution in the world, produced an influential 2012 report on 'Science as an open enterprise'. The Society argued that:

> We are now on the brink of an achievable aim: for all science literature to be online, for all the data to be online and for the two to be interoperable. (Royal Society 2012: 7)

The Society recommended, among other policies, that:

> As a condition of publication, scientific journals should progressively enforce requirements for traceable and usable data... Materials should be uploaded to a repository before the publication of the article (Royal Society 2012: 76)

A large number of journals have adopted such policies; however, the adoption of input sharing policies has been gradual and incomplete. In 2014, Sturges et al. found that only about half of journals surveyed had data sharing policies; a 2013 study of computational research by Stodden et al. found that fewer than 40% of journals had data sharing policies, and only about 20% had code sharing policies. In 2017 Vasilevsky found that only 11.9% of journals surveyed required data sharing as a condition of publication, while 31.8% of journals had no data sharing policy at all. McDonald et al. (2017) found that 63% of observational studies published in the British Medical Journal between 2015 and 2017 "included a statement that implied the data used in the study could not be shared". Similarly, in 2018 Crosas et al. found that only about half of highly-ranked social science journals had data sharing policies, though this figure varied a lot across disciplines. Johnson et al. (2018) found that approximately half of top neurosurgery-related journals had data sharing policies.

Moreover, these policies vary considerably in the degree of input-sharing they request or mandate. Thus for example only 24% of the policies surveyed in Sturges et al. (2014) counted as 'strong' by Piwowar and Chapman's definition[3] (Sturges et al. 2014; Piwowar and Chapman 2008). Similarly, Stodden

[3]Piwowar and Chapman define an input sharing policy as 'weak' if the policy merely suggests that inputs are shared, and 'strong' if it requires "a database accession number as a condition of publication" (Piwowar and Chapman 2008: 3).

et al. found that only about 11% of data- and 3.5% of code-sharing policies required input-sharing as a condition of publication, barring exceptions. (Stodden, Guo, and Ma 2013: 4). Wiley's (2018) survey of engineering journals' data policies found that only four of 28 could be classified as 'strong'.

At the same time, there has been increasing pressure from scholars, advocates and governments for research protocols to be publicly available. As discussed in chapter one, the OpenTrials project aims to establish and maintain:

> an open database of all structured data and documents on all clinical trials, cross-referenced and indexed by trial. (Goldacre and Gray 2016: 2)

This project will aim to reduce the file-drawer and 'researcher degrees of freedom' problems by permitting consumers of research to track the progress of clinical trials from their design to their publication. This in turn seeks, in Goldacre and Gray's words, to:

> drive up standards around open data in evidence-based medicine, and... make different sources more readily comparable and auditable. (Goldacre and Gray 2016: 2)

These incentives are accompanied by the development of alternative metrics for scientific productivity (Priem et al. 2010; Piwowar 2013), including metrics that permit the citation of 'data papers' (Chavan and Penev 2011), code (Mislan, Heer, and White 2016) or other intermediate inputs. These metrics aim to shift the scientific reputational economy away from the Mertonian norm that only the publication of scientific *findings* can result in reputational dividends, towards an alternative reputational economy in which input sharing is more highly valued. Along similar lines, *Psychological Science* has adopted the practice of letting authors "signal open data and materials if they qualified for badges that accompanied published articles" (Kidwell et al. 2016: 1).

Alongside all these shifts, increasing numbers of scientists are adopting open inputs practices as a form of best practice. Indeed, the perception that open inputs are best practice itself functions as an incentive for scientists to participate in the practice (Corti et al. 2019; Easterbrook 2014).

7.5.1 The problem of incentive compatibility

In the chapter so far I have discussed some of the obstacles to implementing open input science, and some of the measures that policymakers and practitioners have taken to try to overcome these obstacles. However, these policies have not always been successful in practice. In a 2014 study of data sharing in economics and management, Andreoli-Versbach and Mueller-Langer found that

> In our sample, 82 empirical researchers (16.8%) sporadically share data, while only 12 (2.46%) share data in a comprehensive and clear way. The large majority, 494 (80.74%), neither share data nor provide any indication of whether or where the data is available. (Andreoli-Versbach and Mueller-Langer 2014)

Along similar lines, a 2017 study of agent-based model code archiving found that only 10% of the articles contained information that allowed readers to locate publicly available source code (Janssen 2017).

In part these shortcomings in open science practice can be attributed to shortcomings of policy. However, as Alsheikh-Ali et al. found in 2011, even when journals adopted input-sharing *policies*, this did not necessarily translate into input-sharing *practices* (Alsheikh-Ali et al. 2011). In their survey, these researchers found that

> Of the remaining 351 papers that were covered by some data availability policy... 208/351 papers (59%) did not adhere fully to the data availability instructions of the journals they were published in (Alsheikh-Ali et al. 2011: 3)

Similarly, Roche et al. looked at 100 papers published in "evolutionary and/or ecological journals" with "strong data archiving policies" and found that:

> In most studies (56%), the archived datasets were incomplete, either because of missing data or insufficient metadata... Secondly, datasets for 64% of studies were archived in a way that either partially or fully prevented reuse... (Roche et al. 2015: 3)

Along similar lines, Van Tuyl and Whitmire (2016) looked at 25 research projects funded by the National Science Foundation and found that, despite data sharing being mandated by the funder's policy and guidelines, 19 (76%) of the projects received a data sharing score of zero. Couture et al. (2018) looked at 315 projects with funder-imposed data publication, in the period 1989-2010, and found that in only 26% of them could data be recovered.

In short, in Savage and Vickers' words:

> our findings suggest that explicit journal policies requiring data sharing do not lead to authors making their data sets available... (Savage and Vickers 2009)

As Chawinga and Zinn (2019), writing a decade later, put it:

> Despite resolute efforts to promote data sharing, relatively little data is shared. (Chawinga and Zinn 2019: 119)

In a recent article that surveyed over 1800 UK-based academics, Zhu (2019) concluded that:

> while most academics recognised the importance of sharing research data, most of them had never shared or reused research data. (Zhu 2019: 1)

These findings imply significant shortcomings in open inputs science as practiced. We can identify two distinct categories of shortcoming in the drive towards open inputs science: on the one hand, the failure of open inputs advocates to implement widespread open inputs policies; on the other hand, the noncompliance by working scientists with those open inputs policies.

This latter problem is a particular concern for open inputs advocates. Increasing the number of journals, institutions, funding bodies, etc. with explicit open inputs policies is a challenge for open inputs advocates, but it is a challenge that can be pursued along similar lines to existing open inputs advocacy. However, most of the *benefits* of open inputs science only accrue if and when inputs are *actually* shared. If open inputs policies do not lead to open inputs *practices*, this raises concerns about the strategies adopted by the open inputs science movement.

This failure can usefully be analysed, I suggest, as a problem of *incentive compatible institution design*. The rules of an institution are 'incentive

incompatible' if the behaviours that the institution or its designers aspire to incentivise are *dominated strategies* in the game-theoretic sense (Börgers 2015). From this theoretical perspective, the goal of institution design - or mechanism design - is to construct institutional incentive structures in which the dominant strategy for all social actors is also one that results in a socially desirable behaviour according to some generally accepted criterion.

From such a perspective, traditional Mertonian science can be seen as a triumph of mechanism design: Mertonian science incentivises the full disclosure of scientific *findings* in a manner that maximises overall social knowledge. Open inputs advocates argue, however, that by *dis*incentivising open input sharing, Mertonian science is incentive *in*compatible with a concept of the general social good that includes broad sharing of intermediate inputs.

Interestingly, the findings of researchers like Alsheikh-Ali et al. (2011) suggest that the incentive structure of Mertonian science is incentive incompatible with open inputs science, even in scenarios in which journals, funding bodies, etc. require intermediate input sharing as a matter of policy. The task for open inputs advocates is therefore not simply to introduce widespread open inputs policy, but to introduce *incentive compatible* open inputs policy. Given the apparent tension between Mertonian and open inputs science, this latter institution-design problem can plausibly be seen as a greater challenge than the former.

In light of this challenge, in the next chapter I extend the formal models of Mertonian science that I developed in Part II, to explore the ways in which incentive compatible open inputs science can be achieved, beginning from a traditional, Mertonian institutional framework.

Chapter 8

Modelling open inputs science

8.1 Introduction

In the previous chapter I discussed debates over open inputs science, and practical challenges faced by policy-makers and practitioners attempting to institute open inputs science. In particular, I outlined two distinct challenges faced by policy-makers aspiring to institutionalise open inputs science. First: the challenge of achieving widespread uptake of open inputs policies. Second: the challenge of designing open inputs policies that are *incentive compatible*. This latter challenge is where the tension between Mertonian science and open inputs science is particularly important. Mertonian science, as I have discussed throughout the thesis, incentivises *open findings* and *closed inputs*, while open inputs science aims to incentivise *both* open findings *and* open inputs. In an institutional situation characterised by scientists' participation in both Mertonian and open inputs incentive structures, there is a tension between closed and open inputs science which different scientists will navigate in different ways. One of the ways in which this tension can be resolved is by scientists using the strategies described in the previous chapter: *partial* sharing of inputs, in a manner that nominally complies with open inputs policies, without the full realisation of open inputs goals. This is the scenario described by many of the empirical studies discussed in section 7.5.1 above.

There is a range of responses to this tension among open inputs advocates. On the one hand, there are those who emphasise the tension between Mertonian and open inputs science. These figures tend to either adopt *pessimistic* attitudes to open inputs science - suggesting that the possibilities of

open inputs science are destined to be unrealised - or *revolutionary* attitudes to Mertonian science - suggesting that a fundamental transformation of scientific institutions is required if open inputs science is to be achieved. On the other hand, there are those who suggest that the tension between Mertonian and open inputs science is overstated, and that the Mertonian norms of 'communism' and of reputational dividends from the sharing of research findings can easily be rerouted into the incentivisation of input sharing and the norm of communally sharing inputs.

These two responses to the apparent tension between Mertonian and open inputs science can be interpreted as a spectrum, with high compatibility at one end of the spectrum and high incompatibility at the other.

In this thesis I adopt a midpoint on this spectrum. I suggest that the tensions between Mertonian and open inputs science are real - but that they are not irreconcilable. To this end, in the sections that follow I modify the models that I developed in Part II of the thesis in ways that remain *within* the Mertonian paradigm, and aim to thereby suggest how open inputs outcomes can be achieved within this paradigm.

In section 8.2, I extend my evolutionary game-theoretic model of scientific cooperation to incorporate a new kind of cooperative strategy: sharing scientific inputs with the entire 'republic of science', not simply the local innovation commons. In section 8.3, I then extend the CODA model of scientific consensus formation to incorporate the possibility of sharing not just findings, but also inputs.

8.2 The prisoner's dilemma of open inputs science

As I discussed in chapter three, within the new institutionalist framework adopted in this thesis there are two broad ways in which economists have analysed institutional transformation: changes in the 'rules of the game' that structure economic life, or changes in the transaction costs that influence institutional decision-making. In this section, I focus on the latter point, and look at the ways in which changes in transactions costs may influence economic decision making within a largely - though not wholly - static scientific action arena.

To do this, I adapt the model of scientific input sharing that I developed

in chapter five, to the problem of *open* input sharing. In chapter five, I looked at the problem of sharing scientific inputs within a network of small, closed, scientific 'innovation commons' - 'invisible colleges' in the sense in which that term is used by Crane, David, and others. I showed that, in a networked scientific community in which there are strong network links *within* subcommunities, but weak network links *between* subcommunities, cooperative behaviour can spread across a network of arbitrarily large size. However, the cooperative behaviour analysed in chapter five was only *locally* cooperative. The problem addressed in this chapter is the problem of 'global' cooperation.

To this end, in the next sub-section I extend the model already developed in chapter five, summarise the results of a series of simulations, and draw some conclusions about the possibilities of incentivising open inputs science.

8.2.1 Extending the evolutionary model

In chapter five, I developed an evolutionary model of traditional Mertonian science, which showed how cooperative behaviour within a specific small-scale innovation commons could spread across a broader scientific network. In that model, I found that strong ties *within* subcommunities and weak ties *between* subcommunities aided the spread of local cooperative behaviour.

However, the cooperative behaviour I studied in that chapter was only cooperative strategies adopted within *small-n* public goods games. The problem of open inputs presents a very different strategic challenge. When sharing open inputs with the entire scientific community - and, moreover, with the public sphere beyond the scientific community - players are engaged in a public goods game with a much larger n.

As I discussed in chapter five, large-n prisoner's dilemmas are both much harder to analyse than small-n dilemmas, and also - intuitively - strategic situations in which it is harder for participants to achieve cooperative outcomes. If we consider an n-player dilemma with a large n simply in its own terms, the challenges to cooperative behaviour seem overwhelming.

This point can be illustrated by considering a model parallel to the one developed in chapter five, in which players engage in an n-player dilemma with all other agents in the model. In this scenario the logic of the prisoner's dilemma implies that cooperation is a strictly dominated strategy, just as it does in a small-n game. In the evolutionary model of chapter five, this logic of the prisoner's dilemma could be overcome, provided some small propor-

tion of the model's sub-communities 'spontaneously' achieved a cooperative equilibrium. These pockets of cooperative behaviour could then 'teach' other subcommunities the benefits that accrue from cooperation, by demonstrating higher payoffs in practice.

This mechanism relies, however, on the initial establishment of cooperative sub-communities. In a scenario in which the typical game involves four players, the probability of any given game achieving the equilibrium of universal cooperation, simply by chance, is $\frac{1}{4!} = \frac{1}{24}$. In a cooperative network composed of significantly more than 24 subcommunities, therefore, the chance of the emergence of some 'pockets' of initial cooperation are high.

In a large-n prisoner's dilemma, this logic does not apply. The chance of a cooperative equilibrium emerging 'naturally' diminishes exponentially as more players are added to the component games. In the limit case in which an entire scientific research field consists of a single game, the chance of a cooperative equilibrium emerging becomes very small indeed.

The comparison of this large-n model and chapter five's small-n model would seem to lead to a pessimistic conclusion for open inputs advocates. These models suggest that, while cooperative behaviour within small-n games can spread across a large graph, provided the graph is structured appropriately, large-n games will 'naturally' remain trapped within the non-cooperative equilibrium of the traditional prisoner's dilemma.

In this chapter, I do not challenge the conclusions drawn from these two models. I suggest, however, that we should not understand open inputs science as a large-n prisoner's dilemma. Instead, I propose that, although open input sharing involves sharing inputs with a large number of other individuals, the game in which this sharing is incentivised may itself usefully be regarded as a small-n game.

Specifically, I extend the model of chapter five by introducing a new strategic option within the small-n games played by invisible colleges or innovation commons. In the model proposed in this chapter, agents can adopt one of *three* strategies. As in the model of chapter five, they can play non-cooperatively, or they can share their inputs with other players in their game. However, in this chapter agents can also share inputs more widely - with the entire public sphere. This decision to broadly share inputs is, however, modelled as a decision made *within* the same small-n public goods game.

How does the strategy to share inputs widely relate to the strategy to share inputs narrowly? In this model, the two strategies are differentiated by a set of parameters. I assume that the costs of open inputs sharing are

higher than the costs of closed inputs sharing. I also assume that there are different reputational payoffs associated with closed and open inputs sharing.

In reputational terms, open data sharing increases the risks to researchers of those outside their collaborative networks freeriding on their research. There are thus significant reputational downsides to sharing data broadly. At the same time, as discussed in section 8.2 above, the reputational payoffs of the scientific game can also be influenced by shifts in academic norms. The reputational benefits of open sharing versus closed sharing may therefore also be *higher* if the community as a whole esteems the practice of open inputs sharing.

In this model the difference between open and closed inputs sharing therefore amounts to differences between the payoffs associated with the two. Relative to closed data sharing, open data sharing increases *transaction costs*, increases the costs of *freeriding*, but also increases the *scientific reputational payoff* for the scientist sharing their inputs. The key question for the relative evolutionary success of these two cooperative strategies is therefore the relative size of these different costs and contributions.

This cost/benefit calculation can be summarised as follows. I define A as overall payoff, R as reputational payoff associated with inputs sharing; T as the transaction cost of input sharing; and F as the reputational costs associated with freeriding. Then

$$A = R - T - F \tag{8.1}$$

We can distinguish these categories of costs and benefits into those associated with closed inputs sharing, and those associated with open inputs sharing, indexed by the subscripts $_c$ and $_o$ respectively. Given the basic framework outlined in chapter five, we would expect open input sharing to become the evolutionarily dominant strategy when the following obtains:

$$A_o > A_c \tag{8.2}$$

That is to say, when

$$(R_o - T_o - F_o) > (R_c - T_c - F_c) \tag{8.3}$$

This conclusion follows intuitively from consideration of the impact of these payoffs on agent behaviour. However, it can also be demonstrated in simulation.

To this end, I ran a series of 220 simulations of 1000 agents interacting for 300 rounds of play. I varied the cost of open inputs sharing, from 2.5 to 3.5, at intervals of 0.1. I ran 20 simulations for each of these values of open input costs, and calculated the mean number of agents adopting each of the three strategies, once play had concluded. The cost of *closed* input sharing remained constant at 3.

Figure 8.1 shows the results of this series of simulations. The system behaves as one would intuitively expect - the strategy which gives the overall highest payoffs is the strategy that disseminates across the network as a whole.

This result is trivial given the basic structure of the model. However, it is not trivial given the problem we started with: the apparent difficulty of incentivising open sharing of scientific inputs, given the size of the scientific community, and the prisoner's dilemma structure of the Mertonian incentive system.

This model shows that one possible solution to the problem of open input sharing, is to reframe this problem at the level of the small-scale scientific innovation commons. If we consider the innovation commons as the locus of the relevant scientific incentive structures and payoffs, then an intractable large-n n-player prisoner's dilemma, becomes a much more tractable small-n n-player prisoner's dilemma. Moreover, as already established in chapter six, the network structure of the 'republic of science' can potentially enable the easy dissemination across the scientific community as a whole, of strategies that are successful at the level of the local innovation commons.

I will return to these findings in the conclusion of the thesis. For now, I turn to the second dimension of the scientific action arena: the problem of epistemic outcomes.

8.3 The opinion dynamics of open inputs science

In the previous section I discussed the possibility of open input sharing spreading across the broader scientific community, using the game-theoretic framework of the n-player prisoner's dilemma. In this section, I will shift focus to the epistemic consequences of open input sharing. As I discussed in chapters one and seven, there are ongoing debates over the strengths and

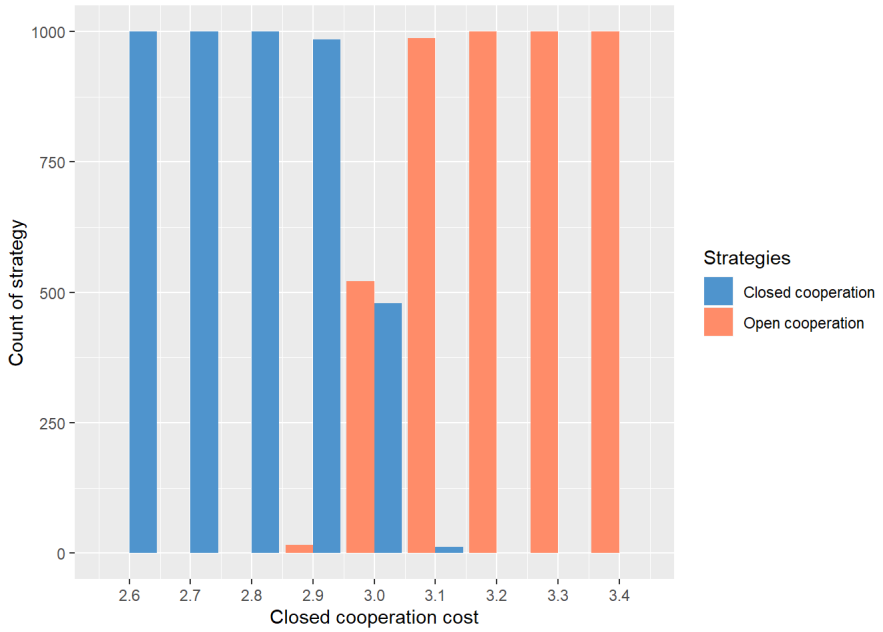


Figure 8.1: Shift from closed to open input sharing, as the relative costs of closed and open cooperation change

weaknesses of open input sharing. In this section I will aim to contribute to those debates by building on the CODA model of scientific opinion dynamics discussed in chapter six.

In section 8.3.1, I review Martins' CODA model, and introduce a variant of the model incorporating open inputs sharing. In section 8.3.2 I run a series of simulations to compare the epistemic outcomes of different degrees of openness in input sharing. I conclude that open input sharing can, in principle, produce a 'step change' in epistemic outcomes, shifting an epistemic system from a 'negative' to a 'positive' equilibrium. However, open inputs can have this positive impact only 'at the margins'. The major determinants of scientific communities' epistemic outcomes remain the frequency and reliability of experimental observation.

8.3.1 Reviewing the CODA model

In chapter six I outlined A.C.R. Martins' continuous opinions, discrete actions (CODA) model of scientific opinion dynamics. The model is characterised by a two-stage decision process for scientific agents. Each agent must make a binary decision about whether or not to accept a scientific proposition P. This decision is informed by a private, subjective probabilistic weighting p of the likelihood that P. Agents can observe nature, or observe the conclusions of their colleagues - but they cannot observe their colleagues' private subjective weightings, or their colleagues' own observations of nature.

In chapter six, I followed Martins in carrying out a series of simulations using the basic CODA model. The challenging scenario, I suggested, is one in which a *false* consensus exists within the scientific community. One of the goals of fallibilist science is to be 'self-correcting': even if a false consensus is established, due to faulty or partial information or reasoning, the scientific research process should be capable of correcting this false consensus over time, as more data is collected, and as the collective process of rational discourse re-evaluates existing conclusions.

This process is, however, mitigated by the potential effects of 'herd behaviour' within scientific opinion dynamics. Once a consensus has been established, this consensus may tend to be 'self-reinforcing', as scientists take their colleagues' beliefs to provide warrant for their own. Even very weak initial evidence can, potentially, snowball into a very strongly held community consensus in favour of false or under-evidenced conclusions.

This is the kind of scenario that concerns many open inputs advocates.

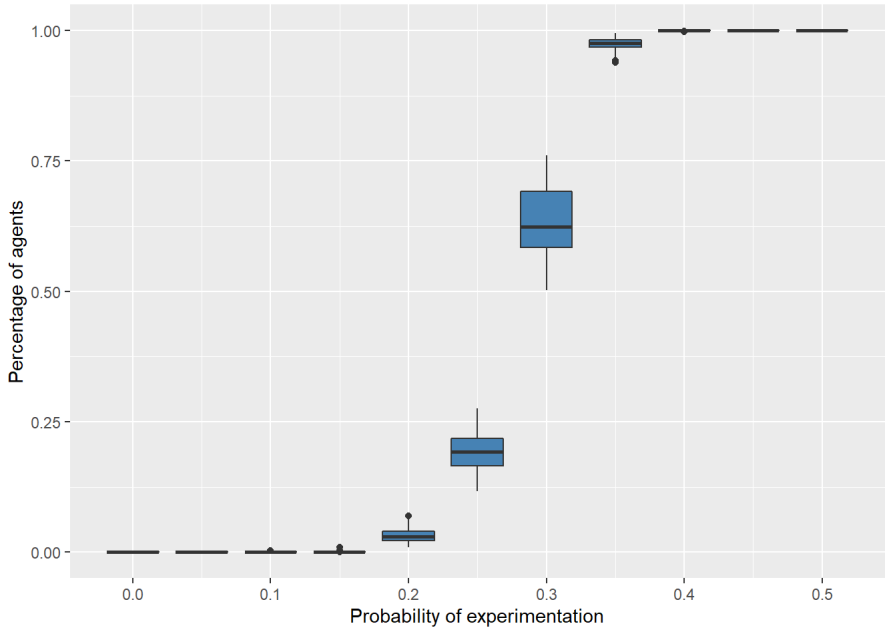


Figure 8.2: CODA model: As the rate of experimentation increases, the percentage of agents believing P changes.

In recent years, as discussed in chapter seven, these concerns have particularly been voiced in connection to the 'replicability crisis' in psychology. The replicability crisis, however, represents only one example of a broader set of concerns that motivate many open inputs institutional proposals. If herd behaviour can potentially lead to an incorrect but entrenched scientific consensus, can the broader sharing of the intermediate inputs of scientific production mitigate this tendency, and lead to better overall epistemic outcomes?

In chapter six, I explored the problem of scientific herd behaviour using Martins' CODA model. For that chapter, I ran a series of simulations, in which an initial false consensus was challenged by observations of nature. Figure 8.2 shows the results of one such set of simulations.

To recapitulate that earlier analysis: in this scenario - starting from weak consensus in favour of falsehood - an experimental rate of more than 15% is required to produce even a weak movement away from false consensus, and an experimental rate of more than 35% to produce a reliable movement towards true consensus. Figure 8.3, by contrast, shows the same scenario,

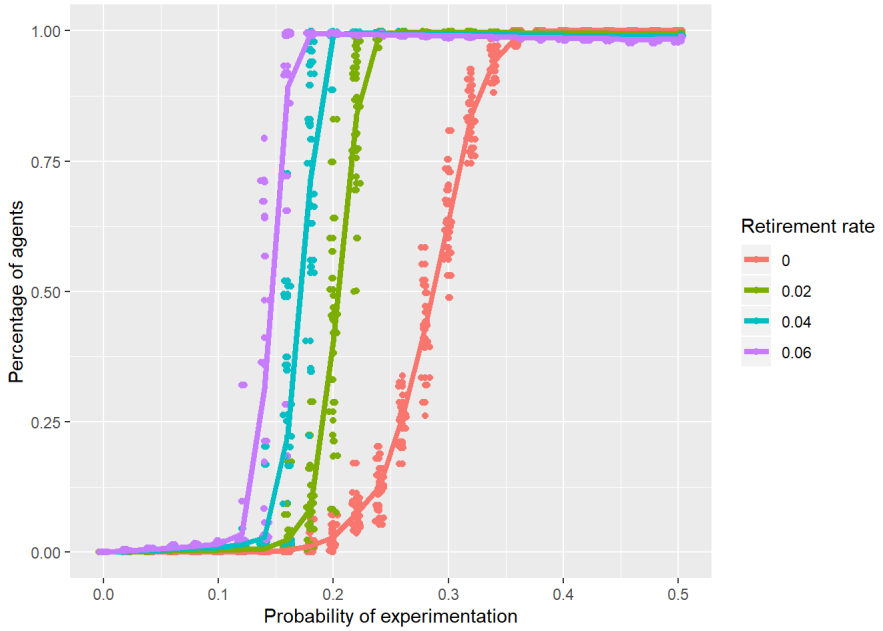


Figure 8.3: CODA model with retirement: Different retirement rates result in different 'tipping points' at which agents shift to believing P.

but with the *retirement* of agents, their replacement by novice researchers, and thereby the regular 'resetting' of agents' priors.

I discuss these findings in more depth in chapter nine. My goal here is to foreground again the 'pessimistic' dimension of these findings. These results suggest that, even in favourable conditions, including weak initial priors and a high turnover of scientists, an initial false consensus will not be supplanted by a true consensus, unless a quite high rate of observation of nature obtains among scientists. This finding is compatible with the scenario pessimistically described by open inputs advocates, in which "most published findings are false" (Ioannidis 2005). From the perspective of open inputs advocates, one worthwhile solution to this problem is the shift from closed to open inputs in scientific practice. In the next section, I therefore introduce open inputs to the model, to evaluate its impact.

8.3.2 Open inputs in the CODA model

In the CODA models analysed so far, agents can individually observe nature, and update their private priors in response to those observations, but can only observe other agents' *conclusions*, rather than their *data* or their *reasoning*. In this section I introduce *open inputs* to the model - the idea that scientists can, in principle, observe not just other scientists' conclusions, but can independently observe records of the observations that lead to these conclusions. If scientists can make such observations of other scientists' inputs, then scientists can also independently evaluate those inputs, on the basis of their own priors, rather than basing their own conclusions so strongly on conclusions partly determined by the priors of others. In principle, this element of open inputs science should reduce the 'herd behaviour' associated with scientific opinion dynamics. In this section I evaluate the consequences of this for our existing CODA model of scientific opinion dynamics.

The open inputs model functions as follows. When an agent observes nature, they keep a record of their observation. This record is retained until the agent's next observation of nature, at which point it is replaced by a record of their new observation. When an agent observes a neighbour, under open inputs science they have a probability k of observing not their neighbour's conclusion, but their neighbour's record of their most recent observation. Thus, in effect, any individual agent has a greater chance of basing their update of their priors on data derived from observation, rather than on another scientist's own conclusions. For working scientists, observing a colleague's openly shared inputs is a more time-consuming process than simply observing a colleague's published findings - hence the fact that scientists will only engage in such 'input checking' in a certain percentage k of interactions. Yet observing another scientist's inputs is also, we can assume, a less costly process than carrying out a new, original experimental observation of nature. Increasing the role of open inputs is therefore, all else equal, a more efficient way to attempt to improve the reliability of the epistemic system's output, than increasing the overall rate of observations.

I ran 20 simulations of 1000 rounds of play each for each of the scenarios under consideration, iterating through different probabilities of experimentation, and rates of open science sharing. Figures 8.4 and 8.5 show the results of these simulations, which each line connecting the mean values of simulation outcomes for a given rate of open science participation.

Figure 8.4 shows the behaviour of the system with an agent replacement

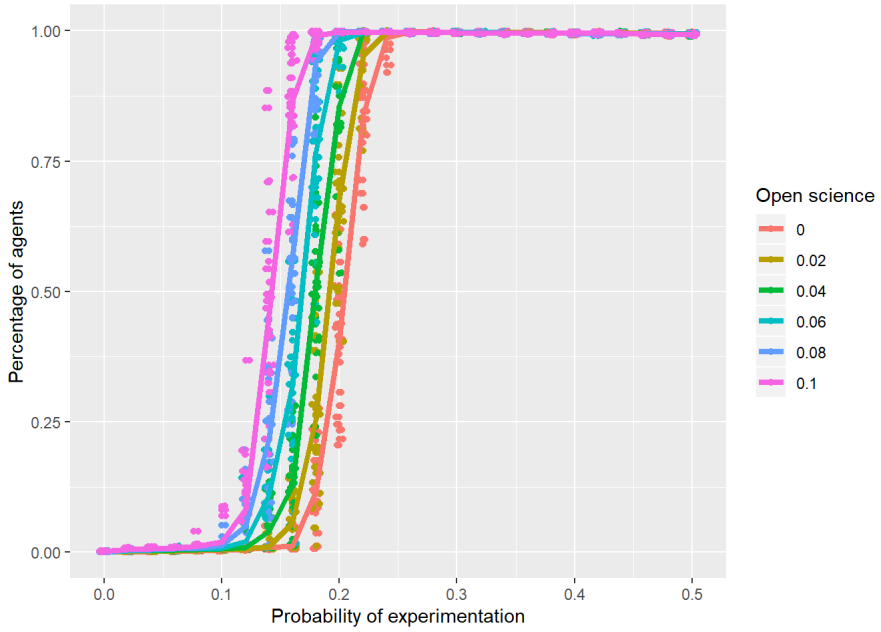


Figure 8.4: CODA model with open inputs science: Each line represents the mean percentage of agents believing P, across a range of different rates of experimentation, for a given rate of open input sharing. This chart represents a rate of agent retirement of 2%

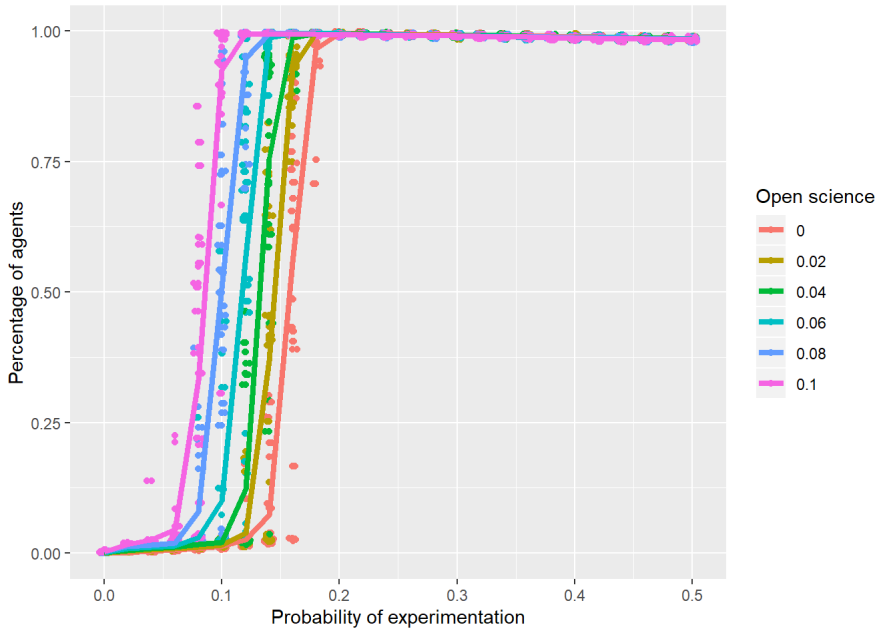


Figure 8.5: CODA model with open inputs science: Agent retirement rate of 5%

rate of 2%. The proportion of agents believing P is again shown as a function of the probability of experimental observations. Each line represents the mean percentage of agents believing P, across a range of different rates of experimentation, for a given rate of open input sharing. Thus the red line represents a probability of open science of 0, the green line a probability of open science of 4%, and so on. Figure 8.5 represents the same scenarios, but with an agent replacement rate of 5%.

As both of these graphs indicate, the impact of open science on the system is - at least given the CODA model's assumptions - as open input advocates suggest. The higher the rate of open inputs observation, the lower the level of experimental observation at which the step change from negative to positive consensus takes place. The 'tipping point' at which a bad consensus shifts to a good consensus moves steadily 'left' across the chart as the rate of open science increases, reducing the number of scenarios in which a bad epistemic consensus is sustained.

8.3.3 Discussion

I am now in a position to assess the impact of open inputs on scientific epistemic outcomes, within the CODA model.

The CODA model usefully represents the mediated structure of scientific opinion-formation. Scientists observe nature, then update their priors about the world based on that observation. But scientists also learn from other scientists, and update their priors based on other scientists' conclusions. The transformation of scientific *inputs* (individual scientists' observations), into scientific *outputs* (individual scientists' findings), and then ultimately into scientific *consensus* (the findings of the scientific community as a whole) thus operates via a web of priors, the dynamics of which can, potentially, override observational inputs in favour of a consensus reproduced purely via (rational, Bayesian) 'herd behaviour'.

In this chapter and chapter six, I have looked at three mechanisms via which the role of herd behaviour in scientific opinion dynamics can be reduced. First, increased observation of nature. Second, retirement of scientists and their replacement by novice researchers with more neutral priors. Third, observation by scientists of their colleagues' *inputs*, rather than simply their colleagues *findings*. All of these changes in scientific practice can reduce the likelihood of 'herd behaviour' false consensus.

One of the conclusions that can be drawn from these results, is that

open inputs themselves, in the absence of a high enough level of repeated high quality experiments, will not result in movement towards a positive equilibrium.

However, the introduction of 'open inputs' to the system of scientific opinion dynamics shifts the point at which these other factors create a 'step change' in consensus behaviour. Thus, for example, a community of scientists with a 15% probability of observing nature in any given round of play, and a 2% replacement rate, will nevertheless remain caught in a 'bad' consensus equilibrium in the absence of open inputs. Introducing a 10% chance of any given scientist observing their neighbour's inputs, however, will reliably shift the system's long-term behaviour from a 'bad' consensus in favour of false belief, to a 'good' consensus in favour of true belief. In this scenario, therefore, open inputs science makes the difference between a consensus in favour of false beliefs, and a consensus in favour of true beliefs.

One of the realistic features of the CODA model, is that a scientific consensus, once established, is 'sticky' due to herd behaviour. This is the reason it is difficult to shift from a bad to a good consensus, even given relatively high rates of observation of nature. However, this feature of the model also means that a good consensus can be sustained once established, with much lower rates of observation of nature.

This suggests that consensus is likely to shift through a sustained period of collective attention to the problem at hand. Conversely, consensus is unlikely to shift through only occasional research, even when correct. A shift to 'open inputs' is therefore not, in itself, sufficient to shift from a bad to a good consensus, even though open inputs can make a significant contribution to such a shift.

This element of the CODA model therefore matches the Kuhnian picture of long periods of 'normal science' driven by background scientific consensus, punctuated by quite rapid transition from one consensus to another. Furthermore, the model suggests that such step change shifts in scientific opinion dynamics are likely to be associated with significant increases in the level of attention paid by scientific researchers to a specific problem-area. Areas of low research interest are unlikely to shift from a bad to a good consensus, even if the research in the area is of high quality.

Nevertheless, the CODA open inputs model proposed in this chapter suggests that open inputs reduce the threshold of collective research effort required, to shift scientific opinion from a bad to a good consensus. In this respect, open input advocates can justifiably claim that institutional shifts

in favour of open inputs can make a significant impact on the epistemic outcomes of science as a whole.

8.4 Summary

In this chapter I have reviewed the agent-based modelling approaches I developed in Part II of the thesis, for the analysis of Mertonian science, and applied these approaches to the analysis of open inputs science. In section 8.2, I discussed the consequences of changing scientific incentive structures for the prisoner's dilemma of cooperative input sharing, and and in section 8.3, I modelled the epistemic consequences of open input sharing. Taken together, these three extensions of my core models enable the comparative analysis of Mertonian and open inputs science.

In the next and final chapter, I will bring the different analyses proposed in the thesis together - to compare Mertonian and open inputs science, and to discuss what lessons can be drawn from this comparison, for the debates over open inputs science.

Chapter 9

Conclusion

The previous chapter concludes the modelling portion of the thesis. I have now developed a series of models - both qualitative and formal - of Mertonian and open inputs science. I have used those models to analyse the behavioural dynamics of scientific *cooperation* and the opinion dynamics of scientific *consensus formation*. In this concluding chapter I will review the argument of the thesis as a whole, and draw some conclusions about the implications of the argument for debates around open inputs science.

9.1 Recapitulation of the argument

I began the thesis, in the Introduction, by summarising the argument of the thesis chapter by chapter. In this conclusion I will review the same argument again, but rather than recapitulate the structure of the thesis, I will now organise my discussion thematically. I will emphasise the three analytic pillars of the argument of the thesis: a *qualitative* account of the structure of science as an institution; a *game-theoretic* account of science as prisoner's dilemma; and an *opinion dynamics* account of science as an epistemic system.

The institutional structure of science

The first pillar of the thesis's argument is a qualitative account of science as an institution. Here I drew on Polanyi's account of 'the republic of science', Merton's account of science as an institution, and Potts' concept of

the innovation commons. In chapter four I argued that the global community of scientific practice is a complex network of partially overlapping subcommunities: a 'republic of science', in Polanyi's sense. Members of this network share specific norms: the Mertonian norms of communism, universalism, disinterestedness, and organised scepticism. They also participate in a shared *reputational economy* in which individual researchers and research teams compete for prestige achieved via *priority of publication*. All else equal, this Mertonian reputational economy incentivises open findings but disincentivises open inputs.

This Polanyian republic of science is internally differentiated into distinct categories of subcommunities - for example, academic disciplines and research networks. In chapter four I argued that one key category of subcommunity is the *innovation commons*: a temporary subcommunity of researchers pooling intellectual resources in the service of a shared research programme. I argued that Ostrom's analysis of commons governance institutions, particularly as extended by researchers affiliated with the Workshop on Governing Knowledge Commons, gives us insight into the norms that govern such scientific innovation commons. At this scale, I argued, it is relatively simple for local innovation commons to institute and enforce norms of input sharing *within* the innovation commons.

This basic picture of science as institution then provided the qualitative framework within which I constructed a series of formal agent-based models of behavioural and opinion dynamics within the scientific community. These formal models focused on two distinct elements of scientific practice: the dynamics of cooperation and noncooperation in input-sharing, and the dynamics of scientific consensus formation. In both of these areas, the thesis was ultimately focused on the differences and continuities between Mertonian and open inputs science.

Science as prisoner's dilemma

The first category of formal model I developed built on Dasgupta and David's account of Mertonian science as n-player prisoner's dilemma. For Dasgupta and David, Mertonian science functions as a prisoner's dilemma because the social good of input sharing is a strictly dominated strategy given the reputational incentive structure associated with priority of publication.

In chapters four and five, I argued that this n-player prisoner's dilemma can be resolved within the *local* scientific innovation commons by means

of conditionally cooperative strategies, and that cooperation *within* innovation commons can spread *between* innovation commons via a form of group selection. Innovation commons that successfully pool intellectual resources are typically more productive than subcommunities that do not - it therefore benefits scientists to participate in input-sharing innovation commons. I modelled the spread of these cooperative strategies across a structured graph, representing a Polanyian 'republic of science'. I demonstrated that cooperative practices spread if there are strong connections *within* and weak connections *between* subcommunities. This finding provides a functionalist explanation for the network structure of the republic of science populated by innovation commons which I described in chapter four.

In chapter eight I then extended this model to the problem of open inputs science. I showed that open inputs practices can *also* spread across a Polanyian network, provided there are greater reputational benefits to researchers for sharing inputs outside the commons than for sharing them within the commons alone. Even in this scenario, I argued, the innovation commons remains an important pivot institution. In chapter four I had proposed a tripartite distinction between the *closed* commons of limited input sharing, the *open* commons of public input sharing, and the *Ostromian* commons of a small-scale scientific governance community. In chapter eight I argued that the institution of Ostromian innovation commons can potentially function as a lever, transforming closed inputs into open inputs science.

This conclusion may not be immediately intuitive, because Ostromian commons are sometimes analysed in ways that suggest they are coterminous with closed commons in the sense of limited resource pooling within a commons as 'club'. There is, however, no intrinsic connection between these different elements of commons governance. One important way forward for open inputs science, I suggest, is using *Ostromian* commons as the basis for *open* commons of intermediate inputs.

Bayesian consensus-formation

The final pillar of the thesis was the formal modelling of scientific opinion-dynamics. Here I built on both Paul David's adaptation of the simple voter model and A.C.R. Martins' more complex 'continuous opinions, discrete actions' CODA model of scientific consensus-formation. The latter model in particular picks out elements of scientific consensus-formation that are highlighted by open science advocates critical of existing scientific institutional

structures. In this CODA model, once the scientific community has formed a consensus around a specific claim, even on the basis of very weak evidence, that consensus can be difficult to shift even when scientists carry out repeated reliable observations of nature that together provide strong evidence against the consensus opinion. This 'stickiness' of scientific opinion is compatible with claims that have emerged from the broad research programme concerned with the 'replicability crisis' in the sciences, particularly the possibility that, in Ioannidis' words, "most published research is false" (Ioannidis 2005).

In chapter eight I added open inputs to this CODA model of scientific consensus-formation. I found that introducing open inputs to this model does not have the revolutionary impact on epistemic outcomes suggested by some of the strongest open inputs advocates; however, open inputs does produce a positive 'step change' in scientific epistemic outcomes in a significant proportion of cases. This thesis therefore adds to the evidence that open inputs science can improve the epistemic outcomes, and thus the reliability and credibility, of scientific research.

Summing up

This recapitulation of the thesis draws together the key strands of my argument. In summing up, I want to highlight four particular contributions the thesis has made to the new economics of science.

First, the qualitative picture of science as institution developed in chapter four brings together a set of intellectual resources that have not to my knowledge hitherto been synthesised. In particular, this thesis represents the first in depth application of the concept of the *innovation commons* to the economics of science.

Second, the game-theoretic models of science as prisoner's dilemma in chapters five and eight present an original account of institutional structures that can foster cooperative practices within a Polanyian scientific network.

Third, the opinion dynamics models of scientific consensus formation in chapters six and eight provide a novel formal account of the dynamics via which scientific input sharing can improve the epistemic outcomes of science as system.

Finally, the thesis brings these elements together into an overarching account of the *relationships* between different elements of scientific institution-design. The contribution of the thesis therefore lies not only in the individual

models presented in each chapter, but also in the way these models combine to create an overall picture of the connections between different elements of the scientific innovation process. In all these ways, the thesis makes an original contribution to the tradition of the new economics of science, building on and going beyond the work of, in particular, Partha Dasgupta and Paul David, Elinor Ostrom, A.C.R. Martins, and the researchers affiliated with the Workshop in Governing Knowledge Commons.

9.2 Implications of the argument

What are the implications of the analysis of the thesis for the debates around open inputs science that I discussed in chapters one and seven? Here I believe three key lessons can be drawn from the analysis presented in the thesis.

First: the challenges of incentive-compatible open inputs institution design can in principle be overcome *within* the overarching institutional structure of traditional Mertonian science. As I discuss in chapter seven, some scholars and advocates have suggested that the tension between Mertonian and open inputs science is a profound one, such that open inputs science can only achieve widespread adoption if fundamental changes to the institutional structures of Mertonian science are also achieved. The thesis presents an argument against this case. I have argued that open inputs science faces a range of institutional challenges, but the incentive structure of Mertonian science is a challenge that can be overcome.

Second: one important *way* in which this challenge can be overcome is via the norm-instituting capacities of relatively small-scale *scientific innovation commons*. In chapter four I argued that the scientific innovation commons is a key scientific institution within which input-sharing can be achieved. In chapter five I argued that the structure of the Polanyian 'republic of science' facilitates the spread of cooperative practices via 'group selection' of the practices adopted by successful innovation commons. Finally, in chapter eight I argued that these features of the scientific innovation commons can be leveraged to facilitate the adoption and spread of *open* inputs practices. Taken together, this argument suggests that the category of the scientific innovation commons could play an important role in the implementation and dissemination of open inputs science.

Third, the thesis makes a contribution to debates over the *epistemic consequences* of open inputs science. In chapters six and eight of the thesis

I developed a model of the opinion dynamics of Mertonian and open inputs science, building on the work of A.C.R. Martins. I found that, in my model, open inputs science *does* improve the epistemic outcomes of scientific research, although other institutional factors - such as the frequency and reliability of observation of nature - are also crucial. The conclusions of the thesis do not support the most hyperbolic claims about the revolutionary impact of open inputs science, but the thesis *does* support the conclusion that open inputs science improves the epistemic outcomes - and thus the reliability and credibility - of scientific knowledge production.

From the perspective of practical implementation of open inputs science by policy-makers and practitioners, I would suggest that the second of these points is the most important. As I discussed in chapter seven, there are a range of institutional levels at which open inputs rules and incentives have been implemented: national and transnational policy bodies; journals and scholarly institutions; institutional repositories; advocacy organisations; etc. The argument of the thesis suggests that an important additional *level* of institutional intervention is the *innovation commons*.

To review the discussion of chapter seven, the goal of open inputs science policy should be to reduce the transaction costs of intermediate input sharing, while increasing the reputational dividend of input sharing. Transaction costs can be reduced through measures such as infrastructure creation, establishment of consensus formats to reduce coordination problems, and reduction in the cost of the relevant information technologies. Reputational dividends can be increased through measures such as journal or funding policies that mandate publication of intermediate inputs, data or code 'papers', professional rewards for the creation of intermediate input resources, and broad acceptance of the community norm of open inputs science, with corresponding local reputational reward and sanction.

These points are aligned with, and derive from, much existing open inputs science policy and research. The argument of this thesis additionally suggests that the *spread* of open input sharing can be achieved via a form of *group selection*. In particular, one key *unit* in this 'evolution of cooperation' is the *innovation commons*.

If this argument is correct, then policymakers and practitioners interested in facilitating the spread of cooperative open inputs practices may be rewarded by focusing on two specific dimensions of scientific institutions.

First, institutional structures that foster 'innovation commons' may be

more likely to result in broad adoption of open inputs science. The Polanyian model of chapter five suggested that scientific research communities with *strong* links between subcommunities - and thus with less internally differentiated subcommunities themselves - are less likely to result in broad cooperative practices spreading across the scientific community as a whole than are research communities that foster relatively self-contained subcommunities of practice. The strong norms of input sharing that open science advocates wish to foster may - paradoxically - be easier to inculcate in relatively 'closed' Ostromian communities, than in more 'open' research communities with 'flat' network structures. This possibility should be considered in debates over 'open innovation', and how best to structure scientific research teams and organisations. It may be that an emphasis on open *networks* may constrain the development of open *commons*.

Second, the argument of the thesis suggests space for an ongoing research programme focused on the principles that inform the successful development of specifically *open inputs* scientific innovation commons. Ostromian analysis in general adopts the attitude that 'one size fits all' approaches to policy are likely to fail, and that policy should be responsive to the local needs of individual commons governance practices. It seems likely that these principles also apply to open science policy, and that, in addition to general incentives to increase the reputational dividends associated with intermediate input sharing, practitioners should focus on the ways in which reputational dividends associated with intermediate input sharing can be increased at the local level of the innovation commons. More work is needed, however, to understand what kinds of reputational incentive structures are most effective at the organisational level of the innovation commons.

9.3 Limitations and opportunities

Finally, I will discuss some additional limitations of the thesis, and opportunities these create for future research in this area.

As I discussed in chapter three, the methodological warrant for the modelling approach of the thesis is what Gibbard and Varian call "caricature". From this methodological perspective, the economic models in the thesis aspire to capture and foreground key elements of social reality, while ignoring others.

This methodological warrant for the approach of the thesis should be

borne in mind, because it means we should be cautious in drawing overly strong conclusions on the basis of the findings. As I discussed in chapters two and three, even within the analytic framework of agent-based modelling of science, a very large number of alternative modelling strategies could potentially be adopted. The models and their findings presented in this thesis should therefore be understood as one possible analytic window on the problems the thesis discusses. There is considerable scope for future work, exploring the implications of taking different approaches to the modelling problems I have tackled in preceding chapters; some of these alternative modelling approaches I intend to pursue in future work.

Alternative pathways for future research suggested by the thesis include, but are not limited to:

- A more nuanced microeconomic account of the rules of the game of scientific innovation commons. This thesis relies on the work of the Workshop on Governing Knowledge Commons as an ongoing empirical project analysing relatively small-scale scientific commons governance communities, of the kind that serve as the 'building blocks' of my game-theoretic model of the 'republic of science'. Yet the problems that preoccupy this thesis are often not the focus of the Workshop researchers. In particular, the shifts in norms, incentives, and 'rules of the game' that can transform a *closed* inputs Ostromian commons into an *open* inputs Ostromian commons are an important topic that merits more empirical research.

- One of the major findings in recent agent-based social epistemology, discussed in chapters two and seven, is Zollman's (2007) argument that greater information sharing between scientists may *reduce* the quality of the epistemic outputs of the system as a whole, because high levels of information sharing may lead to rapid consensus-formation on the basis of inadequate evidence. The CODA model of open inputs science I developed in chapter eight suggests that this consideration may not apply to open inputs science. However, in future work I hope to explore further the question of the circumstances under which Zollman effects do and do not manifest, in relation to open intermediate input sharing.

- One respect in which the models proposed in this thesis are relatively simple is the structure of the graphs within which agents interact. In chapter five these graphs were structured in a manner that varied across

simulations. Nevertheless, this structure remained much simpler and more uniform than real-world scientific networks. Recent work such as Harnagel (2018) has explored the methodological challenges of calibrating agent-based models of science with real-world data about scientific communities, and one possibility for future research would be combining the game-theoretic and opinion dynamics approaches adopted in this thesis with empirically validated graph structures.

- Similarly, although the models in the thesis are dynamic, the graph structures within which agents interact are not. Another possibility for extending the models of the thesis is in developing dynamic graphs, where the links between agents are themselves determined - as they are in real scientific research - by the agents' own choices.

- Finally, the prisoner's dilemma models of cooperation used throughout the thesis are 'symmetrical', in that they assume identical costs and payoffs for all players. As scholars studying open data sharing in the Global South have recently emphasised, however, this assumption often does not hold (Serwadda et al. 2018; Bezuidenhout 2019): some researchers may have greater obstacles to data sharing, and lower expected reputational payoffs, than others, and the distribution of these cost and payoff differentials is non-random across the 'republic of science'. In future work I plan to explore in more depth the implications of this fact for the modelling approaches adopted in this thesis.

These are some of the ways in which the approach I have taken in the thesis could be modified or extended. Naturally there are many more. The field of agent-based modelling of the institutional dimensions of scientific research is an expanding and exciting one. In future research I hope to build on the work in this thesis to look in more depth at many further possibilities for modelling the cooperative and epistemic dimensions of science as an institution.

Bibliography

Abbott, Ryan (2017). "The sentinel initiative as a knowledge commons". In: *Governing medical knowledge commons*. Ed. by Katherine J. Strandburg, Brett M. Frischmann, and Michael J. Madison. Cambridge studies on governing knowledge commons. Cambridge: Cambridge University Press, pp. 121–143.

Acemoglu, Daron, Simon Johnson, and James A. Robinson (2005). "Institutions as a fundamental cause of long-run growth". In: *Handbook of economic growth*. Ed. by Philippe Aghion and Steven N. Durlauf. Vol. 1A. Amsterdam: Elsevier: North-Holland, pp. 385–472.

Acemoglu, Daron and Asuman Ozdaglar (2011). "Opinion dynamics and learning in social networks". In: *Dynamic Games and Applications* 1.1, pp. 3–49.

Acemoglu, Daron and James A. Robinson (2013). *Why nations fail: the origins of power, prosperity and poverty*. London: Profile Books.

Aghion, Philippe and Peter Howitt (1990). *A model of growth through creative destruction*. National Bureau of Economic Research. URL: http://www.nber.org/papers/w3223.

Alexander, Jason McKenzie, Johannes Himmelreich, and Christopher Thompson (2015). "Epistemic landscapes, optimal search, and the division of cognitive labor". In: *Philosophy of Science* 82.3, pp. 424–453.

Aligica, Paul D. and Vlad Tarko (2012). "Polycentricity: From Polanyi to Ostrom, and beyond". In: *Governance* 25.2, pp. 237–262.

Allen, Darcy (2017). "Governing the entrepreneurial discovery of blockchain applications". URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2919170.

Allen, Darcy and Jason Potts (2015). "The innovation commons–why it exists, what it does, who it benefits, and how". URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2617141.

Allen, Darcy and Jason Potts (2016). "How innovation commons contribute to discovering and developing new technologies". In: *International Journal of the Commons* 10.2. URL: https://www.thecommonsjournal.org/articles/10.18352/ijc.644/?toggle_hypothesis=on.

Alsheikh-Ali, Alawi A. et al. (2011). "Public availability of published research data in high-impact journals". In: *PloS one* 6.9, e24357. URL: https://doi.org/10.1371/journal.pone.0024357.

Ames, E (1961). "Research, invention, development and innovation". In: *The American Economic Review* 51.3, pp. 370–381.

Andreoli-Versbach, Patrick and Frank Mueller-Langer (2014). "Open access to data: An ideal professed but not practised". In: *Research Policy* 43.9, pp. 1621–1633.

Aoki, Masahiko (2001). *Toward a comparative institutional analysis*. Cambridge, MA: MIT Press.

ARDC (2019). *Australian Research Data Commons*. URL: https://ardc.edu.au/ (visited on 01/19/2019).

Arrow, Kenneth J. (1962). "The economic implications of learning by doing". In: *The Review of Economic Studies* 29.3, pp. 155–173.

Attard, Judie et al. (2015). "A systematic review of open government data initiatives". In: *Government Information Quarterly* 32.4, pp. 399–418.

Avin, Shahar (2017). "Centralised funding and epistemic exploration". URL: http://philsci-archive.pitt.edu/13331/.

Axelrod, Robert and William D. Hamilton (1981). "The evolution of cooperation". In: *Science* 211.4489, pp. 1390–1396.

Axelrod, Robert M. (2006). *The evolution of cooperation: Revised edition*. New York: Basic Books.

Ayres, Clarence Edwin (1946). *The theory of economic progress*. Chapel Hill: University of North Carolina Press.

Balconi, Margherita, Stefano Brusoni, and Luigi Orsenigo (2010). "In defence of the linear model: An essay". In: *Research Policy* 39.1, pp. 1–13.

Ball, Philip (2015). "Leading mathematician launches arXiv 'overlay' journal". In: *Nature News* 526.7571, p. 146.

Ballot, Gerard, Antoine Mandel, and Annick Vignes (2015). "Agent-based modeling and economic theory: Where do we stand?" In: *Journal of Economic Interaction and Coordination* 10.2, pp. 199–220.

Barber, Bernard (1952). *Science and the social order*. New York: Free Press.

Bartling, Sönke and Sascha Friesike (2014a). *Opening science: The evolving guide on how the internet is changing research, collaboration and scholarly publishing*. Berlin: Springer.

— (2014b). "Towards another scientific revolution: The evolving guide on how the internet is changing research, collaboration and scholarly publishing". In: *Opening Science*. Ed. by Sönke Bartling and Sascha Friesike. Berlin: Springer, pp. 3–15.

Ben-Naim, Eli (2005). "Opinion dynamics: Rise and fall of political parties". In: *Europhysics Letters* 69.5, p. 671.

Benkler, Yochai (2002). "Coase's penguin, or, Linux and 'The nature of the firm'". In: *The Yale Law Journal* 112.3, pp. 369–446.

— (2006). *The wealth of networks: How social production transforms markets and freedom*. New Haven: Yale University Press.

Berger, Peter L. and Thomas Luckmann (1991). *The social construction of reality: A treatise in the sociology of knowledge*. London: Penguin UK.

Berman, Elizabeth Popp (2011). *Creating the market university: How academic science became an economic engine*. Princeton, NJ: Princeton University Press.

Bernal, John Desmond (1939). *The social function of science*. London: George Routledge & Sons, Ltd.

Bezuidenhout, Louise (2019). "To share or not to share: Incentivizing data sharing in life science communities". In: *Developing World Bioethics* 19.1, pp. 18–24.

Bianchi, Federico and Flaminio Squazzoni (2015). "Is three better than one? Simulating the effect of reviewer selection and behavior on the quality and efficiency of peer review". In: *2015 Winter simulation conference (WSC)*. IEEE, pp. 4081–4089.

Björk, Bo-Christer et al. (2010). "Open access to the scientific journal literature: Situation 2009". In: *PLoS ONE* 5.6, e11273. URL: `http://dx.plos.org/10.1371/journal.pone.0011273`.

Björk, Bo-Christer et al. (2014). "Anatomy of green open access". In: *Journal of the Association for Information Science and Technology* 65.2, pp. 237–250.

Bloor, David (1976). *Knowledge and social imagery*. London: Routledge and Kegan Paul.

— (2011). *The enigma of the aerofoil: Rival theories in aerodynamics, 1909-1930*. Chicago, Ill: University of Chicago Press.

Bogers, Marcel, Henry Chesbrough, and Carlos Moedas (2018). "Open innovation: Research, practices, and policies". In: *California Management Review* 60.2, pp. 5–16.

Boggio, Andrea (2017). "Population biobanks' governance: A case study of knowledge commons". In: *Governing medical knowledge commons*. Ed. by Brett M. Frischmann, Katherine J. Strandburg, and Michael J. Madison. Cambridge: Cambridge University Press, pp. 102–120.

Bolick, Cheryl Mason (2006). "Digital archives: Democratizing the doing of history". In: *International Journal of Social Education* 21.1, pp. 122–134.

Bonilla, Jesús P. Zamora (1999). "The elementary economics of scientific consensus". In: *Theoria: An International Journal for Theory, History and Foundations of Science* 14.3, pp. 461–488.

— (2002). "Scientific inference and the pursuit of fame: A contractarian approach". In: *Philosophy of Science* 69.2, pp. 300–323.

— (2012). "The economics of scientific knowledge". In: *Handbook of the philosophy of science: Philosophy of economics*. Ed. by Uskali Mäki, pp. 823–862.

Bonney, Rick et al. (2014). "Next steps for citizen science". In: *Science* 343.6178, pp. 1436–1437.

Borg, Anne Marie et al. (2018). "Epistemic effects of scientific interaction: Approaching the question with an argumentative agent-based model". In: *Historical Social Research/Historische Sozialforschung* 43.1, pp. 285–307.

Boyle, James (2003a). "Foreword: The opposite of property?" In: *Law and Contemporary Problems* 66, pp. 1–32.

— (2003b). "The second enclosure movement and the construction of the public domain". In: *Law and Contemporary Problems* 66, pp. 33–74.

Brandom, Robert (1994). *Making it explicit: Reasoning, representing, and discursive commitment*. Cambridge, Mass; London: Harvard University Press.

— (2002). *Tales of the mighty dead: Historical essays in the metaphysics of intentionality*. Cambridge, MA: Harvard University Press.

— (2008). *Between saying and doing: Towards an analytic pragmatism*. Oxford: Oxford University Press.

— (2009a). *Articulating reasons*. Cambridge, MA: Harvard University Press.

— (2009b). *Reason in philosophy: Animating ideas*. Cambridge, MA: Harvard University Press.

Brase, Jan (2014). "Making data citeable: DataCite". In: *Opening science: The evolving guide on how the internet is changing research, collabora-*

tion and scholarly publishing. Ed. by Sönke Bartling and Sascha Friesike. Berlin: Springer, pp. 327–329.

Börgers, Tilman (2015). *An introduction to the theory of mechanism design*. New York: Oxford University Press, USA.

Buchanan, JM (1965). "An economic theory of clubs". In: *Economica* 32.125, pp. 1–14.

Burks, Arthur W. and John Von Neumann (1966). *Theory of self-reproducing automata*. Champaign, IL: University of Illinois Press.

Bush, Vannevar (1945). *Science: The endless frontier*. Washington, DC: United States Government Printing Office.

Callon, Michel (1994). "Is science a public good?" In: *Science, Technology, & Human Values* 19.4, pp. 395–424.

Camerer, Colin F. et al. (2016). "Evaluating replicability of laboratory experiments in economics". In: *Science* 351.6280, pp. 1433–1436.

Campbell, Donald T. (1960). "Blind variation and selective retentions in creative thought as in other knowledge processes." In: *Psychological Review* 67.6, pp. 380–400.

Cartwright, Nancy (2010). "Models: Parables v fables". In: *Beyond mimesis and convention: Representation in art and science*. Ed. by Roman Frigg and Matthew Hunter. Dordrecht Heidelberg London New York: Springer, pp. 19–31.

Chalmers, David J. (2006). "Strong and weak emergence". In: *The re-emergence of emergence*. Ed. by P. Clayton and P. Davies. Oxford: Oxford University Press, pp. 244–256.

Chan, Leslie et al. (2002). *Budapest open access initiative*. URL: `http://www.budapestopenaccessinitiative.org/read`.

Chaudhuri, Ananish (2011). "Sustaining cooperation in laboratory public goods experiments: A selective survey of the literature". In: *Experimental Economics* 14.1, pp. 47–83.

Chavalarias, David (2017). "What's wrong with science?" In: *Scientometrics* 110.1, pp. 481–503.

Chavan, Vishwas and Lyubomir Penev (2011). "The data paper: A mechanism to incentivize data publishing in biodiversity science". In: *BMC Bioinformatics* 12.15, S2.

Chawinga, Winner Dominic and Sandy Zinn (2019). "Global perspectives of research data sharing: A systematic literature review". In: *Library & Information Science Research* 41.2, pp. 109–122.

Chesbrough, Henry, Wim Vanhaverbeke, and Joel West (2006). *Open innovation: Researching a new paradigm*. Oxford: Oxford University Press.
— (2014). *New frontiers in open innovation*. Oxford: Oxford University Press.
Chong, Sei and Dennis Normile (2006). "How young Korean researchers helped unearth a scandal..." In: *Science* 311.5757, pp. 22–25.
Clark, John Maurice (1939). *Social control of business*. New York, London: McGraw-Hill.
Clifford, Peter and Aidan Sudbury (1973). "A model for spatial conflict". In: *Biometrika* 60.3, pp. 581–588.
Coase, Ronald H. (1937). "The nature of the firm". In: *Economica* 4.16, pp. 386–405.
— (1960). "The problem of social cost". In: *The Journal of Law and Economics* 3, pp. 1–44.
— (1998). "The new institutional economics". In: *The American Economic Review* 88.2, pp. 72–74.
CoCoRaHS (2019). *About CoCoRaHS*. URL: `http://www.cocorahs.org/Content.aspx?page=aboutus` (visited on 10/09/2019).
Cohen, GA (2000). *Karl Marx's thery of history: A defence*. Oxford: Clarendon Press.
Collins, H. M. and Graham Cox (1976). "Recovering relativity: Did prophecy fail?" In: *Social Studies of Science* 6.3, pp. 423–444.
Collins, Harry M. (1974). "The TEA set: Tacit knowledge and scientific networks". In: *Science Studies* 4.2, pp. 165–185.
— (1983). "An empirical relativist programme in the sociology of scientific knowledge". In: *Science observed: Perspectives on the social study of science*. Ed. by Karin Knorr-Cetina and Michael Mulkay. London and Beverley Hills: Sage, pp. 85–113.
Commons, John R. (1931). "Institutional economics". In: *The American Economic Review* 21.4, pp. 648–657.
— (1932). "The problem of correlating law economics and ethics". In: *Wisconsin Law Review* 8, pp. 3–26.
— (1934). *Institutional economics: Its place in political economy*. New York: Macmillan.
Contreras, Jorge L. (2014). "Constructing the genome commons". In: *Governing knowledge commons*. Ed. by Katherine J. Strandburg, Michael J. Madison, and Brett M. Frischmann. Oxford: Oxford University Press, pp. 112–13.

Cooke, Philip, Mikel G. Uranga, and Goio Etxebarria (1998). "Regional systems of innovation: An evolutionary perspective". In: *Environment and Planning A: Economy and Space* 30.9, pp. 1563–1584.

Cooter, Robert D. (1989). "The Coase Theorem". In: *Allocation, Information and Markets*. Ed. by John Eatwell, Murray Milgate, and Peter Newman. London: Macmillan, pp. 64–70.

Corti, Louise et al. (2019). *Managing and sharing research data: A guide to good practice*. 2nd. Google-Books-ID: PkSxDwAAQBAJ. SAGE. 369 pp.

Couture, Jessica L. et al. (2018). "A funder-imposed data publication requirement seldom inspired data sharing". In: *PloS one* 13.7, e0199789.

Crane, Diana (1972). *Invisible colleges: Diffusion of knowledge in scientific communities*. Chicago, Ill: University of Chicago Press.

Crea, Robert (2004). *DNA chemistry at the dawn of commercial biotechnology*.

Crokidakis, Nuno, Victor H. Blanco, and Celia Anteneodo (2014). "Impact of contrarians and intransigents in a kinetic model of opinion dynamics". In: *Physical Review E* 89.1, p. 013310.

Crosas, Mercè et al. (2018). "Data policies of highly-ranked social science journals". In:

Dasgupta, Partha (1989). "The economics of parallel research". In: *The economics of missing markets, information, and games*. Ed. by Frank Hahn. Oxford: Clarendon Press, pp. 129–48.

Dasgupta, Partha and Paul A. David (1987). "Information disclosure and the economics of science and technology". In: *Arrow and the ascent of modern economic theory*. Ed. by G. Feiwel. New York: New York University Press. (Visited on 08/16/2015).

— (1994). "Toward a new economics of science". In: *Research Policy* 23.5, pp. 487–521.

David, Paul A. (1994). "Positive feedbacks and research productivity in science: Reopening another black box". In: *Economics and technology*. Ed. by O Grandstrand. Amsterdam: Elsevier.

— (1998a). "Common agency contracting and the emergence of "open science" institutions". In: *The American Economic Review* 88.2, pp. 15–21.

— (1998b). "Communication norms and the collective cognitive performance of "invisible colleges"". In: *Creation and transfer of knowledge*. Ed. by Giorgio Barba Navaretti et al. Berlin Heidelberg: Springer, pp. 115–163.

Davidson, Sinclair and Jason Potts (2016). "A new institutional approach to innovation policy". In: *Australian Economic Review* 49.2, pp. 200–207.

De Langhe, Rogier (2014). "A unified model of the division of cognitive labor". In: *Philosophy of Science* 81.3, pp. 444–459.

DeGroot, Morris H. (1974). "Reaching a consensus". In: *Journal of the American Statistical Association* 69.345, pp. 118–121.

Djankov, Simeon et al. (2003). "The new comparative economics". In: *Journal of Comparative Economics* 31.4, pp. 595–619.

DOAJ (2019). *Directory of Open Access Journals*. URL: `https://doaj.org` (visited on 09/15/2019).

Domar, Evsey D. (1946). "Capital expansion, rate of growth, and employment". In: *Econometrica* 14.2, pp. 137–147.

Dung, Phan Minh (1995). "On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games". In: *Artificial Intelligence* 77.2, pp. 321–357.

Eamon, William (1985). "From the secrets of nature to public knowledge: The origins of the concept of openness in science". In: *Minerva* 23.3, pp. 321–347.

Earp, Brian D. and David Trafimow (2015). "Replication, falsification, and the crisis of confidence in social psychology". In: *Frontiers in Psychology* 6, p. 621.

Easterbrook, Steve M. (2014). "Open code for open science?" In: *Nature Geoscience* 7.11, p. 779.

eBird (2019). *About eBird*. URL: `http://ebird.org/content/ebird/about/` (visited on 10/20/2019).

Edgerton, David (2004). "'The linear model' did not exist: Reflections on the history and historiography of science and research in the twentieth century". In: *The science-industry nexus: History, policy, implications*. Ed. by Karl Grandin, Nina Wormbs, and Sven Widmalm. New York: Watson.

Epstein, Joshua M. and Robert Axtell (1996). *Growing artificial societies: Social science from the bottom up*. Washington, DC: Brookings Institution Press.

Etzkowitz, Henry and Loet Leydesdorff (2000). "The dynamics of innovation: from National Systems and "Mode 2" to a Triple Helix of university–industry–government relations". In: *Research Policy* 29.2, pp. 109–123.

European Commission (2016). *Open innovation, open science, open to the world - a vision for Europe*. URL: `https://ec.europa.eu/digital-`

single-market/en/news/open-innovation-open-science-open-world.

European Commission (2017). *EOSC declaration*. URL: https://ec.europa.eu/research/openscience/pdf/eosc_declaration.pdf.

— (2018). *Prompting an EOSC in practice: Final report and recommendations of the Commission 2nd High Level Expert Group on the European Open Science Cloud*. URL: https://publications.europa.eu/en/publication-detail/-/publication/5253a1af-ee10-11e8-b690-01aa75ed71a1.

Everett, Jim Albert Charlton and Brian D. Earp (2015). "A tragedy of the (academic) commons: Interpreting the replication crisis in psychology as a social dilemma for early-career researchers". In: *Frontiers in Psychology* 6, p. 1152.

Farmer, J. Doyne and Duncan Foley (2009). "The economy needs agent-based modelling". In: *Nature* 460.7256, p. 685.

Fecher, Benedikt and Sascha Friesike (2014). "Open science: One term, five schools of thought". In: *Opening science*. New York: Springer, pp. 17–47.

Fecher, Benedikt, Sascha Friesike, and Marcel Hebing (2015). "What drives academic data sharing?" In: *PloS one* 10.2, e0118053.

Fennell, Lee Anne (2011). "Ostrom's law: Property rights in the commons". In: *International Journal of the Commons* 5.1, pp. 9–27. URL: https://www.thecommonsjournal.org/articles/10.18352/ijc.252/.

Franke, Nikolaus and Frank Piller (2004). "Value creation by toolkits for user innovation and design: The case of the watch market". In: *The Journal of Product Innovation Management* 21.6, pp. 401–415.

Franke, Nikolaus and Martin Schreier (2002). "Entrepreneurial opportunities with toolkits for user innovation and design". In: *International Journal on Media Management* 4.4, pp. 225–234.

Frankenhuis, Wilem and Daniel Nettle (2018). "Open science is liberating and can foster creativity". In: *Perspectives on Psychological Science* 13.4, pp. 439–447.

Freeman, Christopher (1987). *Technology policy and economic performance*. London; New York: Frances Pinter Publishers.

— (1995). "The 'National System of Innovation' in historical perspective". In: *Cambridge Journal of Economics* 19.1, pp. 5–24.

Freeman, Richard B. (2011). "The economics of science and technology policy". In: *The science of science policy: A handbook*. Ed. by Kaye Husbands Fealing et al. Stanford, CA: Stanford University Press, pp. 85–103.

Freese, Jeremy and David Peterson (2017). "Replication in social science". In: *Annual Review of Sociology* 43, pp. 147–165.

Friedman, JW (1971). "A non-cooperative equilibirum for supergames". In: *The Review of Economic Studies* 38.1, pp. 1–12.

Friedman, Milton (1953). "The Methodology of Positive Economics". In: *Essays in positive economics*. Chicago, Ill: University of Chicago Press, pp. 3–43.

Frischmann, Brett M., Michael J. Madison, and Katherine J. Strandburg (2014). *Governing knowledge commons*. Oxford: Oxford University Press.

Fudenberg, Drew and Eric Maskin (1986). "The folk theorem in repeated games with discounting or with incomplete information". In: *Econometrica: Journal of the Econometric Society* 54.3, pp. 533–554.

Furnas, Clifford Cook (1948). *Research in industry: Its organization and management*. New York: D. Van Nostrand Co.

Furubotn, Eirik Grundtvig and Rudolf Richter (2005). *Institutions and economic theory: The contribution of the new institutional economics*. Ann Arbor: University of Michigan Press.

Galam, Serge (2002). "Minority opinion spreading in random geometry". In: *The European Physical Journal B-Condensed Matter and Complex Systems* 25.4, pp. 403–406.

Galam, Serge, Yuval Gefen, and Yonathan Shapir (1982). "Sociophysics: A new approach of sociological collective behaviour. I. mean-behaviour description of a strike". In: *The Journal of Mathematical Sociology* 9.1, pp. 1–13.

Gardner, Paul A. (1995). "The relationship between technology and science: Some historical and philosophical reflections. Part II". In: *International Journal of Technology and Design Education* 5.1, pp. 1–33.

Garfinkle, Harold (1967). *Studies in ethnomethodology*. Englewood Cliffs: Prentice-Hall.

Gassmann, Oliver, Ellen Enkel, and Henry Chesbrough (2010). "The future of open innovation". In: *R&D Management* 40.3, pp. 213–221.

Geertz, Clifford (1980). "Blurred genres: The refiguration of social thought". In: *The American Scholar* 49.2, pp. 165–179.

Gelman, Andrew and Eric Loken (2013). "The garden of forking paths: Why multiple comparisons can be a problem, even when there is no "fishing expedition" or "p-hacking" and the research hypothesis was posited ahead of time". In: *Department of Statistics, Columbia University*. URL: `http:`

//www.stat.columbia.edu/~gelman/research/unpublished/p_hacking.pdf.
Gewin, Virginia (2012). "Uncovering misconduct". In: *Nature* 485.7396, pp. 137–139.
Gezelter, Dan (2011). *An informal definition of OpenScience \textbackslash textbar The OpenScience Project*. URL: http://www.openscience.org/blog/?p=454 (visited on 08/07/2015).
Gibbard, Allan and Hal R. Varian (1978). "Economic models". In: *The Journal of Philosophy* 75.11, pp. 664–677.
Gibbons, Michael et al. (1994). *The new production of knowledge: The dynamics of science and research in contemporary societies*. London: Sage.
Gieryn, Thomas F. (1983). "Boundary-work and the demarcation of science from non-science: Strains and interests in professional ideologies of scientists". In: *American sociological review* 48.6, pp. 781–795.
— (1999). *Cultural boundaries of science: Credibility on the line*. Chicago, Ill: University of Chicago Press.
Gilbert, Nigel (1997). "A simulation of the structure of academic science". In: *Sociological Research Online* 2.2, pp. 1–15.
Gilboa, Itzhak et al. (2014). "Economic models as analogies". In: *The Economic Journal* 124.578, F513–F533.
Ginsparg, Paul (1997). "Winners and losers in the global research village". In: *The Serials Librarian* 30.3, pp. 83–95.
Glanvill, Joseph (1668). *Plus ultra, or, The progress and advancement of knowledge since the days of Aristotle in an account of some of the most remarkable late improvements of practical, useful learning, to encourage philosophical endeavours: occasioned by a conference with one of the notional way*. London: James Collins.
Godin, Benoît (2006). "The linear model of innovation: The historical construction of an analytical framework". In: *Science, Technology, & Human Values* 31.6, pp. 639–667.
Goldacre, Ben and Jonathan Gray (2016). "OpenTrials: Towards a collaborative open database of all available information on all clinical trials". In: *Trials* 17.1, p. 164.
Goldman, Alvin I. (1999). *Knowledge in a social world*. Oxford: Oxford University Press.
Goldman, Alvin I. and Moshe Shaked (1991). "An economic model of scientific activity and truth acquisition". In: *Philosophical Studies* 63.1, pp. 31–55.

González, Andrés Guadamuz (2005). "Open science: Open source licenses in scientific research". In: *North Carolina Journal of Law and Technology* 7.2, pp. 321–366.

Gotts, Nicholas Mark, J. Gareth Polhill, and Alistair N. R. Law (2003). "Agent-based simulation in the study of social dilemmas". In: *Artificial Intelligence Review* 19.1, pp. 3–92.

Greif, Avner (1998). "Historical and comparative institutional analysis". In: *The American Economic Review* 88.2, pp. 80–84.

Grim, Patrick (2009). "Threshold phenomena in epistemic networks". In: *2009 AAAI Fall Symposium Series*.

Hacking, Ian (1999). *The social construction of what?* Cambridge, MA: Harvard University Press.

Hamill, Lynne and G. Nigel Gilbert (2016). *Agent-based modelling in economics*. Wiley Online Library.

Hamilton, Walton H. (1919). "The institutional approach to economic theory". In: *The American Economic Review* 9.1, pp. 309–318.

Hands, D. Wade (1994a). "Blurred boundaries: Recent changes in the relationship between economics and the philosophy of natural science". In: *Studies in History and Philosophy of Science Part A* 25.5, pp. 751–772.

— (1994b). "The sociology of scientific knowledge: Some thoughts on the possibilities". In: *New directions in economic methodology*. Ed. by Roger E. Backhouse. London; New York: Routledge, pp. 75–108.

— (1997). "Conjectures and reputations: The sociology of scientific knowledge and the history of economic thought". In: *History of Political Economy* 29.4, pp. 695–739.

Haraway, Donna (1988). "Situated knowledges: The science question in feminism and the privilege of partial perspective". In: *Feminist studies* 14.3, pp. 575–599.

Hardin, Garrett (1968). "The tragedy of the commons". In: *Science* 162.3859, pp. 1243–1248.

Harding, Sandra (1986). *The science question in feminism*. Ithaca; London: Cornell University Press.

— (1993). *The racial economy of science: Toward a democratic future*. Bloomington, IN: Indiana University Press.

Hardwicke, Tom E. et al. (2018). "Data availability, reusability, and analytic reproducibility: Evaluating the impact of a mandatory open data policy at the journal Cognition". In: *Royal Society Open Science* 5.8, p. 180448. (Visited on 10/03/2019).

Hardwig, John (1985). "Epistemic dependence". In: *The Journal of Philosophy* 82.7, pp. 335–349.

Harnagel, Audrey (2018). "A mid-level approach to modeling scientific communities". In: *Studies in History and Philosophy of Science Part A* In press.

Harrod, Roy F. (1939). "An essay in dynamic theory". In: *The Economic Journal* 49.193, pp. 14–33.

Hausman, Daniel M. (1992). *The inexact and separate science of economics*. Cambridge: Cambridge University Press.

Hayek, Friedrich August (1945). "The use of knowledge in society". In: *The American Economic Review* 35.4, pp. 519–30.

Hegselmann, Rainer, Ulrich Krause, and others (2002). "Opinion dynamics and bounded confidence: Models, analysis, and simulation". In: *Journal of Artificial Societies and Social Simulation* 5.3. URL: `http://jasss.soc.surrey.ac.uk/5/3/2.html`.

Henrich, Joseph Patrick (2004). *Foundations of human sociality: Economic experiments and ethnographic evidence from fifteen small-scale societies*. Oxford: Oxford University Press. 472 pp.

Hess, Charlotte and Elinor Ostrom (2007). *Understanding knowledge as a commons*. Cambridge, MA: The MIT Press.

Highfill, Jannett K. (1992). "Economics as allegory". In: *Journal of Interdisciplinary Economics* 4.2, pp. 131–136.

Hirsch, Paul, Stuart Michaels, and Ray Friedman (1987). "Dirty hands" versus "clean models". In: *Theory and Society* 16.3, pp. 317–336.

Hodgson, Geoffrey M. (1993). "Institutional economics: Surveying the 'old' and the 'new'". In: *Metroeconomica* 44.1, pp. 1–28.

— (2004). *The evolution of institutional economics*. London: Routledge.

Holley, Richard A. and Thomas M. Liggett (1975). "Ergodic theorems for weakly interacting infinite systems and the voter model". In: *The Annals of Probability* 3.4, pp. 643–663.

Huijboom, Noor and Tijs Van den Broek (2011). "Open data: An international comparison of strategies". In: *European Journal of ePractice* 12.1, pp. 4–16.

Hull, David L. (1988). *Science as a process: An evolutionary account of the social and conceptual development of science*. Chicago, Ill: University of Chicago Press.

Hunter, Michael (1981). *Science and society in Restoration England*. Cambridge: Cambridge University Press.

Hunter, Michael (1989). *Establishing the new science: The experience of the early Royal Society*. Woodbridge, Suffolk: Boydell & Brewer Ltd.

— (1994). *The Royal Society and its Fellows 1660-1700: The morphology of an early scientific institution*. Oxford: The Alden Press.

Hutchison, Terence W. (1984). "Institutionalist economics old and new". In: *Zeitschrift für die gesamte Staatswissenschaft/Journal of Institutional and Theoretical Economics* 140 (H. 1), pp. 20–29.

Ioannidis, John PA (2005). "Why most published research findings are false". In: *PLoS medicine* 2.8, e124.

Irwin, Alan (1995). *Citizen science: A study of people, expertise and sustainable development*. London: Routledge. (Visited on 08/21/2015).

Jackson, Matthew O. (2010). *Social and economic networks*. Princeton, NJ: Princeton University Press.

Janssen, Marco A. (2017). "The practice of archiving model code of agent-based models". In: *Journal of Artificial Societies and Social Simulation* 20.1. URL: `http://jasss.soc.surrey.ac.uk/20/1/2.html`.

Jasanoff, Sheila (2004). *States of knowledge: The co-production of science and the social order*. London: Routledge.

Jefferson, Thomas (1854). *The writings of Thomas Jefferson: Correspondence*. Vol. VI. New York: Riker, Thorne and Co.

John, Leslie K., George Loewenstein, and Drazen Prelec (2012). "Measuring the prevalence of questionable research practices with incentives for truth telling". In: *Psychological Science* 23.5, pp. 524–532.

Johnson, Jeremiah N. et al. (2018). "Data sharing in neurosurgery and neurology journals". In: *Cureus* 10.5, e2680.

Kahn, Jonathan (1999). "Biotechnology and the legal constitution of the self: Managing identity in science, the market, and society." In: *Hasting Law Journal* 51, pp. 909–952.

Kawrykow, Alexander et al. (2012). "Phylo: A citizen science approach for improving multiple sequence alignment". In: *PLoS ONE* 7.3, e31362.

Kerr, Norbert L. (1998). "HARKing: Hypothesizing after the results are known". In: *Personality and Social Psychology Review* 2.3, pp. 196–217.

Khatib, Firas et al. (2011). "Algorithm discovery by protein folding game players". In: *Proceedings of the National Academy of Sciences of the United States of America* 108.47, pp. 18949–18953.

Kidwell, Mallory C. et al. (2016). "Badges to acknowledge open practices: A simple, low-cost, effective method for increasing transparency". In: *PLoS Biology* 14.5, e1002456.

Kitcher, Philip (1993). *The advancement of science: Science without legend, objectivity without illusions*. Oxford: Oxford University Press.

Klein, Peter G. (2000). "New institutional economics". In: URL: `https://ssrn.com/abstract=115811`.

Kline, Stephen J. and others (1985). "Innovation is not a linear process". In: *Research Management* 28.4, pp. 36–45.

Kline, Stephen J. and Nathan Rosenberg (1986). "An overview of innovation". In: *The positive sum strategy: Harnessing technology for economic growth*. Ed. by Ralph Landau and Nathan Rosenberg. Washington, DC: National Academy Press, pp. 275–305.

Kovanis, Michail et al. (2017). "Evaluating alternative systems of peer review: A large-scale agent-based modelling approach to scientific publication". In: *Scientometrics* 113.1, pp. 651–671.

Kreps, David M. and Robert Wilson (1982). "Reputation and imperfect information". In: *Journal of Economic Theory* 27.2, pp. 253–279.

Krige, John (2006). "Critical reflections on the science-technology relationship". In: *Transactions of the Newcomen Society* 76.2, pp. 259–269.

Kuhn, T. (1962). *The structure of scientific revolutions*. Chicago, Ill: University of Chicago Press.

Kummerfeld, Erich and Kevin JS Zollman (2015). "Conservatism and the scientific state of nature". In: *The British Journal for the Philosophy of Science* 67.4, pp. 1057–1076.

Lakatos, Imre and others (1970). *Criticism and the growth of knowledge: Volume 4: Proceedings of the International Colloquium in the Philosophy of Science, London, 1965*. Cambridge: Cambridge University Press.

Larson, Maja and Margaret Chon (2017). "The greatest generational Impact: Open neuroscience as an emerging knowledge commons". In: *Governing medical knowledge commons*. Ed. by Brett M. Frischmann, Katherine J. Strandburg, and Michael J. Madison. Cambridge: Cambridge University Press, pp. 166–191.

Laudan, Larry (1981). "The pseudo-science of science?" In: *Philosophy of the Social Sciences* 11, pp. 173–98.

Law, Duncan (2014). "Scientific boundary-work and the contested status of open science". In: TASA 2014: Challenging Identities, Institutions and Communities. Adelaide: The Australian Sociological Association (TASA).

— (2016). "Theorising internal institutional diversity: Elinor Ostrom on rules and social conflict". In: *Refereed Proceedings of TASA 2016 Conference*.

TASA 2016: Cities and Successful Societies. Melbourne: The Australian Sociological Association (TASA), pp. 206–212.

Law, Duncan and Nicole Pepperell (2014). "Knowledge and social practice: Robert Brandom and the strong programme". In: *Refereed Proceedings of TASA 2014 Conference*. TASA 2014: Challenging Identities, Institutions and Communities. Adelaide: The Australian Sociological Association (TASA), pp. 1–19.

— (2017). "Sociology and the mirror of nature: Robert Brandom and the strong programme". In: *Journal of Sociology* 53.1, pp. 245–263.

— (2018). "Oppression in the commons: Cautionary notes on Elinor Ostrom's concept of self-governance". In: *Refereed Proceedings of TASA 2018 Conference*. TASA 2018: Precarity, Rights and Resistance. Ed. by Grazyna Zajdow. Melbourne: The Australian Sociological Association (TASA), pp. 7–12.

Lerner, Josh and Jean Tirole (2000). "The simple economics of open source". In: *NBER Working Paper* 7600.

Lessig, Lawrence (2002). *The future of ideas: The fate of the commons in a connected world*. New York: Vintage.

— (2004). "The creative commons". In: *Montana Law Review* 65.1, pp. 1–13.

— (2009). "The Tragedy of the Innovation Commons: Reconciling Private Claims with Public Interest". In: *JL Phil. & Culture* 4, p. 35.

Lessig, Lawrence and Nico Stehr (2004). "The innovation commons". In: *The governance of knowledge*. Piscataway, NJ: Transaction Publishers, pp. 241–254.

Levine, Peter (2007). "Collective action, civic engagement, and the knowledge commons". In: *Understanding knowledge as a commons*. Ed. by Charlotte Hess and Elinor Ostrom. Cambridge, MA: MIT Press, pp. 247–276.

Lewis, David (2008). *Convention: A philosophical study*. Oxford: Blackwell Publishers.

Lewontin, Richard C. (1961). "Evolution and the theory of games". In: *Journal of theoretical biology* 1.3, pp. 382–403.

Lin, Songqing (2013). "Why serious academic fraud occurs in China". In: *Learned Publishing* 26.1, pp. 24–27.

Lintott, Chris J. et al. (2008). "Galaxy Zoo: Morphologies derived from visual inspection of galaxies from the Sloan Digital Sky Survey". In: *Monthly Notices of the Royal Astronomical Society* 389.3, pp. 1179–1189.

Lintott, Chris J. et al. (2009). "Galaxy Zoo: 'Hanny's Voorwerp', a quasar light echo?" In: *Monthly Notices of the Royal Astronomical Society* 399.1, pp. 129–140.

Loewenstein, George F., Leigh Thompson, and Max H. Bazerman (1989). "Social utility and decision making in interpersonal contexts." In: *Journal of Personality and Social Psychology* 57.3, p. 426.

Lotka, Alfred J. (1926). "The frequency distribution of scientific productivity". In: *Journal of the Washington Academy of Sciences* 16.12, pp. 317–323.

Lucas, Robert E. (1988). "On the mechanics of economic development". In: *Journal of Monetary Economics* 22.1, pp. 3–42.

Lucey, Morgan, Jocalyn Clark, and Paul Glasziou (2017). "Public availability of trial protocols". In: *The Lancet* 390.10113, e54–e55.

Maclaurin, W. Rupert (1953). "The sequence from invention to innovation and its relation to economic growth". In: *The Quarterly Journal of Economics* 67.1, pp. 97–111.

Madison, Michael J. (2014). "Commons at the intersection of peer production, citizen science, and big data: Galaxy Zoo". In: *Governing knowledge commons*. Ed. by Brett M. Frischmann, Michael J. Madison, and Katherine J. Strandburg, pp. 209–254.

Madison, Michael J., Brett M. Frischmann, and Katherine J. Strandburg (2009a). "Constructing commons in the cultural environment". In: *Cornell Law Review* 95.4, pp. 657–709.

— (2009b). "The university as constructed cultural commons". In: *Washington University Journal of Law and Policy* 30, pp. 365–403.

Madison, Michael J., Katherine J. Strandburg, and Brett M. Frischmann (2016). "Knowledge commons". In: *Research handbook on the economics of intellectual property law (Vol. II - Analytical methods)*. Ed. by Peter Menell and David Scwartz. Cheltenham: Edward Elgar Publishing.

Maki, Uskali (2002). *Fact and fiction in economics: Models, realism and social construction*. Cambridge: Cambridge University Press.

Martins, André CR (2008). "Continuous opinions and discrete actions in opinion dynamics problems". In: *International Journal of Modern Physics C* 19.4, pp. 617–624.

— (2009). "Bayesian updating rules in continuous opinion dynamics models". In: *Journal of Statistical Mechanics: Theory and Experiment* 2009.2, P02017.

Martins, André CR (2010). "Modeling scientific agents for a better science". In: *Advances in Complex Systems* 13.4, pp. 519–533.

Martins, André CR, Carlos de B. Pereira, and Renato Vicente (2009). "An opinion dynamics model for the diffusion of innovations". In: *Physica A: Statistical Mechanics and its Applications* 388.15, pp. 3225–3232.

Martins, T. Vaz, Miguel Pineda, and Raul Toral (2010). "Mass media and repulsive interactions in continuous-opinion dynamics". In: *Europhysics Letters* 91.4, p. 48003.

Masuzzo, Paola and Lennart Martens (2017). *Do you speak open science? Resources and tips to learn the language*. PeerJ Preprints. URL: `https://peerj.com/preprints/2689/`.

McCloskey, Deirdre N. (1998). *The rhetoric of economics*. Madison, Wisconsin: University of Wisconsin Press.

McDonald, Laura et al. (2017). "A review of data sharing statements in observational studies published in the BMJ: A cross-sectional study". In: *F1000Research* 6.1708.

McGucken, William (1978). "On freedom and planning in science: The society for freedom in science, 1940-46". In: *Minerva* 16.1, pp. 42–72.

Merton, Robert K. (1957a). "Priorities in scientific discovery: A chapter in the sociology of science". In: *American Sociological Review* 22.6, pp. 635–659.

— (1957b). *Social theory and social structure*. Glencoe, Ill.: Free Press.

— (1961). "Singletons and multiples in scientific discovery: A chapter in the sociology of science". In: *Proceedings of the American Philosophical Society* 105.5, pp. 470–486.

— (1968). "The Matthew effect in science". In: *Science* 159.3810, pp. 56–63.

— (1973). "The normative structure of science". In: *The sociology of science: Theoretical and empirical investigations*. Ed. by Norman W. Storer. Chicago, Ill: University of Chicago Press, pp. 267–278.

— (1988). "The Matthew effect in science, II: Cumulative advantage and the symbolism of intellectual property". In: *Isis* 79.4, pp. 606–623.

Mill, John Stuart (1877). *Essays on some unsettled questions of political economy*. London: John W. Parker.

Mills, Charles (2007). "White ignorance". In: *Race and epistemologies of ignorance* 247, pp. 26–31.

Mirowski, Philip (2011). *Science-mart*. Cambridge, MA: Harvard University Press.

Mislan, K. A. S., Jeffrey M. Heer, and Ethan P. White (2016). "Elevating the status of code in ecology". In: *Trends in Ecology & Evolution* 31.1, pp. 4–7.

Mitchell, Wesley Clair (1913). *Business cycles*. Berkeley, California: University of California Press.

— (1927). *Business cycles: The problem and its setting*. New York: National Bureau of Economic Research.

— (1959). *Business cycles and their causes*. Berkeley, California: University of California Press.

Mobilia, Mauro, A. Petersen, and Sidney Redner (2007). "On the role of zealotry in the voter model". In: *Journal of Statistical Mechanics: Theory and Experiment* 2007.8, P08029.

Mokyr, Joel (2016). *A culture of growth: The origins of the modern economy*. Princeton, NJ: Princeton University Press.

Molloy, Jennifer C. (2011). "The open knowledge foundation: Open data means better science". In: *PLoS Biology* 9.12, e1001195. URL: `http://dx.plos.org/10.1371/journal.pbio.1001195`.

Mons, Barend (2018). *Data stewardship for open science: Implementing FAIR principles*. Boca Raton, FL: CRC Press.

Mons, Barend et al. (2017). "Cloudy, increasingly FAIR: Revisiting the FAIR Data guiding principles for the European Open Science Cloud". In: *Information Services & Use* 37.1, pp. 49–56.

Morgan, Mary S. (2001). "Models, stories and the economic world". In: *Journal of Economic Methodology* 8.3, pp. 361–384.

Moyle, Martin and Antony Lewis (2008). *RIOJA (Repository Interface to Overlaid Journal Archives) project*. URL: `http://discovery.ucl.ac.uk/12562/`.

Muldoon, Ryan and Michael Weisberg (2011). "Robustness and idealization in models of cognitive labor". In: *Synthese* 183.2, pp. 161–174.

Munafo, Marcus R. et al. (2017). "A manifesto for reproducible science". In: *Nature Human Behaviour* 1.1, pp. 1–9. (Visited on 10/03/2019).

National Academies of Sciences and Medicine (2018). *Open science by design: Realizing a vision for 21st century research*. Washington, DC: National Academies Press.

National Institutes of Health (2018). "NIH strategic plan for data science". In: *Office of Data Science Strategy*. URL: `https://datascience.nih.gov/strategicplan`.

Naughton, Linda and David Kernohan (2016). "Making sense of journal research data policies". In: *Insights* 29.1.

NBN (2019). *The national biodiversity network*. URL: http://www.nbn.org.uk/The-NBN/What-is-the-NBN.aspx.

Nelson, Andrew J. (2016). "How to share "a really good secret": Managing sharing/secrecy tensions around scientific knowledge disclosure". In: *Organization Science* 27.2, pp. 265–285.

Nelson, Richard R. (1959). "The simple economics of basic scientific research". In: *Journal of Political Economy* 67.3, pp. 297–306.

— (1962). "The link between science and invention: The case of the transistor". In: *The rate and direction of inventive activity: Economic and social factors*. Ed. by National Bureau of Economic Research. Princeton, New Jersey: Princeton University Press, pp. 549–584.

— (2004). "The market economy, and the scientific commons". In: *Research Policy* 33.3, pp. 455–471.

Neylon, Cameron and Shirley Wu (2009). "Open science: Tools, approaches, and implications". In: *Pacific Symposium on Biocomputing 2009*. Ed. by Russ B. Altman et al., pp. 540–544.

Nielsen, Michael (2012). *Reinventing discovery: the new era of networked science*. Princeton and Oxford: Princeton University Press.

NIH (2018). *NIH strategic plan for data science*. URL: https://commonfund.nih.gov/sites/default/files/NIH_Strategic_Plan_for_Data_Science_Final_508.pdf.

North, Douglass C. (1990). *Institutions, institutional change and economic performance*. Cambridge: Cambridge University Press.

North, Douglass C., John Joseph Wallis, and Barry R. Weingast (2006). *A conceptual framework for interpreting recorded human history*.

— (2009). *Violence and social orders: A conceptual framework for interpreting recorded human history*. Cambridge: Cambridge University Press.

Nosek, Brian A. et al. (2015). "Promoting an open research culture". In: *Science* 348.6242, pp. 1422–1425.

Nosek, Brian A. et al. (2018). "The preregistration revolution". In: *Proceedings of the National Academy of Sciences* 115.11, pp. 2600–2606.

Nowotny, Helga, Peter Scott, and Michael Gibbons (2003). "Introduction: Mode 2 revisited: The new production of knowledge". In: *Minerva* 41.3, pp. 179–194.

Nugroho, Rininta Putri et al. (2015). "A comparison of national open data policies: Lessons learned". In: *Transforming Government: People, Process and Policy* 9.3, pp. 286–308.

Olle, Meritxell Ramirez-i (2015). "Rhetorical strategies for scientific authority: A boundary-work analysis of 'climategate'". In: *Science as Culture* 24.4, pp. 384–411.

Olson, Mancur (1965). *The logic of collective action: Public goods and the theory of groups*. Cambridge, MA: Harvard University Press.

Open Research Data Task Force (2018). *Realising the potential: Final report of the Open Research Data Task Force*. URL: `https://www.gov.uk/government/publications/open-research-data-task-force-final-report`.

Open Science Collaboration (2015). "Estimating the reproducibility of psychological science". In: *Science* 349.6251, aac4716.

OpenContent (1999). *Open publication license*. URL: `http://www.opencontent.org/openpub/`.

O'Riordan, Colm and Humphrey Sorensen (2008). "Stable cooperation in the n-player prisoner's dilemma: The importance of community structure". In: *Adaptive Agents and Multi-Agent Systems III. Adaptation and Multi-Agent Learning*. Ed. by K. Tuyls et al. Verlag Berlin Heidelberg: Springer, pp. 157–168.

Ostrom, Elinor (1990). *Governing the commons: The evolution of institutions for collective action*. Cambridge: Cambridge University Press.

— (2005). *Understanding institutional diversity*. Princeton, NJ: Princeton University Press.

Pepperell, Nicole and Duncan Law (2015). "The Internet imaginary: Between technology and technique". In: *M/C Journal* 18.2. URL: `http://journal.media-culture.org.au/index.php/mcjournal/article/view/957`.

Perc, Matjaž et al. (2013). "Evolutionary dynamics of group interactions on structured populations: A review". In: *Journal of the Royal Society Interface* 10.80, p. 20120997.

Phillips, Melissa Lee (2006). "Do you need an electronic lab notebook?" In: *The Scientist* 20.3, pp. 74–75.

Pickering, Andrew (1992). "From science as knowledge to science as practice". In: *Science as practice and culture*. Ed. by Andrew Pickering. Chicago; London: University of Chicago Press, pp. 1–26.

Piwowar, Heather et al. (2018). "The state of OA: A large-scale analysis of the prevalence and impact of Open Access articles". In: *PeerJ* 6, e4375.

Piwowar, Heather A. (2013). "Altmetrics: Value all research products". In: *Nature* 493.7431, p. 159.
Piwowar, Heather A. and Wendy Webber Chapman (2008). "A review of journal policies for sharing research data." In: *AMIA Annual Symposium Proceedings 2008*, pp. 1–14.
Planck, Max and Rogers D. Rusk (1950). "Scientific autobiography and other papers". In: *American Journal of Physics* 18.2, pp. 117–117.
Polanyi, Michael (1941). "The growth of thought in society". In: *Economica* 8.32, pp. 428–456.
— (1945). "The autonomy of science". In: *The Scientific Monthly* 60.2, pp. 141–150.
— (1958). *Personal knowledge: Towards a post-critical philosophy*. Chicago, Ill: University of Chicago Press.
— (1967). *The tacit dimension*. London: Routledge and Kegan Paul.
— (2000). "The republic of science: Its political and economic theory". In: *Minerva* 38.1, pp. 1–32.
Popper, Karl R. (1959). *The logic of scientific discovery*. London: Hutchinson.
— (1963). *Conjectures and refutations*. London: Routledge.
— (1979). *Objective knowledge: An evolutionary approach*. Oxford: Oxford University Press.
— (2002). *The poverty of historicism*. London: Routledge.
Potts, Jason (2012). *The innovation commons*. URL: `https://ssrn.com/abstract=2706856`.
— (2018). "Governing the innovation commons". In: *Journal of Institutional Economics* 14.6, pp. 1025–1047.
Potts, Jason and John Hartley (2015). "How the social economy produces innovation". In: *Review of Social Economy* 73.3, pp. 263–282.
Priem, Jason and Bradley M. Hemminger (2012). "Decoupling the scholarly journal". In: *Frontiers in Computational Neuroscience* 6.19.
Priem, Jason et al. (2010). *Altmetrics: A manifesto*. URL: `http://altmetrics.org/manifesto`.
Proskurnikov, Anton V., Alexey S. Matveev, and Ming Cao (2016). "Opinion dynamics in social networks with hostile camps: Consensus vs. polarization". In: *IEEE Transactions on Automatic Control* 61.6, pp. 1524–1536.
Purver, Margery (2013). *The Royal Society: Concept and creation*. London: Routledge.
Pöyhönen, Samuli (2017). "Value of cognitive diversity in science". In: *Synthese* 194.11, pp. 4519–4540.

Quattrociocchi, Walter, Guido Caldarelli, and Antonio Scala (2014). "Opinion dynamics on interacting networks: Media competition and social influence". In: *Scientific Reports* 4, p. 4938.

Quine, Willard V. (1951). "Two dogmas of empiricism". In: *The Philosophical Review* 60.1, pp. 20–43.

Rabin, Matthew (1993). "Incorporating fairness into game theory and economics". In: *The American Economic Review* 83.5, pp. 1281–1302.

Raddick, M. Jordan et al. (2010). "Galaxy zoo: Exploring the motivations of citizen science volunteers". In: *Astronomy Education Review* 9.1.

Rezaei, Golriz, Michael Kirley, and Jens Pfau (2009). "Evolving cooperation in the n-player prisoner's dilemma: A social network model". In: *Artificial life: Borrowing from biology*. 4th Australian Conference on Artificial Life, 2009. Ed. by Kevin Korb, Marcus Randall, and Tim Hendtlass. Berlin: Springer, pp. 43–52.

Righi, Simone and Károly Takács (2017). "The miracle of peer review and development in science: An agent-based model". In: *Scientometrics* 113.1, pp. 587–607.

Riordan, Michael H. and Oliver E. Williamson (1985). "Asset specificity and economic organization". In: *International Journal of Industrial Organization* 3.4, pp. 365–378.

Roche, Dominique G. et al. (2015). "Public data archiving in ecology and evolution: How well are we doing?" In: *PLoS Biology* 13.11, e1002295.

Rodriguez, Marko A., Johan Bollen, and Herbert Van de Sompel (2006). "The convergence of digital libraries and the peer-review process". In: *Journal of Information Science* 32.2, pp. 149–159.

Romer, Paul M. (1986). "Increasing returns and long-run growth". In: *Journal of Political Economy* 94.5, pp. 1002–1037.

— (1990). "Endogenous technological change". In: *Journal of Political Economy* 98.5, S71–S102.

Rosenthal, Robert (1979). "The file drawer problem and tolerance for null results." In: *Psychological Bulletin* 86.3, p. 638.

Royal Society (2012). *Science as an open enterprise*. The Royal Society Science Policy Centre. URL: `https://royalsociety.org/topics-policy/projects/science-public-enterprise/report/`.

Rubinstein, Ariel (2006). "Dilemmas of an economic theorist". In: *Econometrica* 74.4, pp. 865–883.

— (2012). *Economic fables*. Cambridge: Open Book Publishers.

Rutherford, Malcolm (1996). *Institutions in economics: The old and the new institutionalism*. Cambridge: Cambridge University Press.
— (2000). "Understanding institutional economics: 1918-1929". In: *Journal of the History of Economic Thought* 22.3, pp. 277–308.
Samuelson, Paul A. (1954). "The pure theory of public expenditure". In: *The Review of Economics and Statistics* 36.4, pp. 387–389.
Savage, Caroline J. and Andrew J. Vickers (2009). "Empirical study of data sharing by authors publishing in PLoS journals". In: *PloS one* 4.9, e7078.
Schelling, Thomas (1978). *Micromotives and macrobehavior*. New York: WW Norton.
Schelling, Thomas C. (1969). "Models of segregation". In: *The American Economic Review* 59.2, pp. 488–493.
Schumpeter, Joseph A. (1994). *Capitalism, socialism and democracy*. London: Routledge.
— (2017). *Theory of economic development*. London: Routledge.
Science International (2015). *Open data in a big data world*. URL: `https://council.science/cms/2017/04/open-data-in-big-data-world_long.pdf`.
Searle, John R. (1995). *The construction of social reality*. New York: Simon and Schuster.
Sen, Amartya (1981). *Poverty and famines: An essay on entitlement and deprivation*. Oxford: Oxford University Press.
Serwadda, David et al. (2018). "Open data sharing and the Global South—Who benefits?" In: *Science* 359.6376, pp. 642–643.
Shapin, Steven and Simon Schaffer (1985). *Leviathan and the air-pump: Hobbes, Boyle, and the experimental life*. Princeton, NJ: Princeton University Press.
Shrout, Patrick E. and Joseph L. Rodgers (2018). "Psychology, science, and knowledge construction: Broadening perspectives from the replication crisis". In: *Annual Review of Psychology* 69, pp. 487–510.
Silvertown, Jonathan (2009). "A new dawn for citizen science". In: *Trends in Ecology & Evolution* 24.9, pp. 467–471.
Simmonds, Rob et al. (2016). "The African data intensive research cloud". In: *IST-Africa Week Conference, 2016*. IEEE, pp. 1–8.
Simmons, Joseph P., Leif D. Nelson, and Uri Simonsohn (2011). "False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant". In: *Psychological Science* 22.11, pp. 1359–1366.

Siva, Nayanah (2014). "Crowdfunding for medical research picks up pace". In: *The Lancet* 384.9948, pp. 1085–1086.

Smaldino, Paul E. and Richard McElreath (2016). "The natural selection of bad science". In: *Royal Society Open Science* 3.9, p. 160384.

Smaldino, Paul E., Matthew A. Turner, and Pablo A. Contreras Kallens (2019). "Open science and modified funding lotteries can impede the natural selection of bad science". In: *Royal Society Open Science* 6.7, p. 190194.

Smith, Adam (1976). *The wealth of nations*. Chicago, Ill: University of Chicago Press.

Smith, John Maynard (1972). *On evolution*. Edinburgh: Edinburgh University Press.

— (1988). "Evolution and the theory of games". In: *Did Darwin get it right?* Ed. by John Maynard Smith. Berlin: Springer, pp. 202–215.

Smith, John Maynard and George R. Price (1973). "The logic of animal conflict". In: *Nature* 246.5427, pp. 15–18.

Smith, John WT (1999). "The deconstructed journal: A new model for academic publishing". In: *Learned Publishing* 12.2, pp. 79–91.

Sobkowicz, Pawel (2015). "Innovation suppression and clique evolution in peer-review-based, competitive research funding systems: An agent-based model". In: *Journal of Artificial Societies and Social Simulation* 18.2, p. 13.

Sokal, Alan and Jean Bricmont (1999). *Fashionable nonsense: Postmodern intellectuals' abuse of science*. New York: Picador.

Solow, Robert M. (1956). "A contribution to the theory of economic growth". In: *The Quarterly Journal of Economics* 70.1, pp. 65–94.

— (1957). "Technical change and the aggregate production function". In: *The Review of Economics and Statistics* 39.3, pp. 312–320.

— (1994). "Perspectives on growth theory". In: *The Journal of Economic Perspectives* 8.1, pp. 45–54.

Sîrbu, Alina et al. (2017). "Opinion dynamics: Models, extensions and external effects". In: *Participatory sensing, opinions and collective awareness*. Ed. by Vittorio Loreto et al. Berlin: Springer, pp. 363–401.

Stafford, Ned (2010). "Science in the digital age". In: *Nature* 467.7317, S19–S21.

Stallman, Richard (2015). *What is free software?* URL: `http://www.gnu.org/philosophy/free-sw.html` (visited on 09/16/2015).

Stauffer, Dietrich and Muhammad Sahimi (2007). "Can a few fanatics influence the opinion of a large segment of a society?" In: *The European Physical Journal B* 57.2, pp. 147–152.

Stevens, Raymond (1941). *A report on industrial research as a national resource: Introduction*. United States National Resources Committee.

Stewart, Walter W. et al. (1919). "Economic theory: Discussion". In: *The American Economic Review* 9.1, pp. 319–324.

Stiglitz, Joseph E. (1999). "Knowledge as a global public good". In: *Global public goods: International cooperation in the 21st century*. Ed. by Inge Kaul, Isabelle Grunberg, and Marc A. Stern. New York; Oxford: Oxford University Press, pp. 308–25.

Stodden, Victoria, Peixuan Guo, and Zhaokun Ma (2013). "Toward reproducible computational research: An empirical analysis of data and code policy adoption by journals". In: *PloS one* 8.6, e67111.

Stokes, Donald E. (1994). "Completing the Bush model: Pasteur's quadrant". In: *Science The Endless Frontier: Learning from the Past, Designing for the Future. Highlights from the Conference Series*, pp. 23–35.

— (1997). *Pasteur's quadrant: Basic science and technological innovation*. Washington, DC: Brookings Institution Press.

Strandburg, Katherine, Brett Frischmann, and Can Cui (2014). "The Rare Diseases Clinical Research Network and the Urea Cycle Disorders Consortium as nested knowledge commons". In: *Governing knowledge commons*. Ed. by Brett M. Frischmann, Michael J. Madison, and Katherine J. Strandburg. Oxford: Oxford University Press, pp. 155–208.

Strandburg, Katherine J., Brett M. Frischmann, and Michael J. Madison (2017). "The knowledge commons framework". In: *Governing medical knowledge commons*. Ed. by Brett M. Frischmann, Katherine J. Strandburg, and Michael J. Madison. Cambridge: Cambridge University Press, pp. 9–18. URL: https://doi.org/10.1017/9781316544587.

Stroebe, Wolfgang, Tom Postmes, and Russell Spears (2012). "Scientific misconduct and the myth of self-correction in science". In: *Perspectives on Psychological Science* 7.6, pp. 670–688.

Sturges, Paul et al. (2014). "Access to research data: Addressing the problem through journal data sharing policies". In: IATUL Conference 2014.

Suber, Peter (2015). *Open access overview*. URL: http://legacy.earlham.edu/peters/fos/overview.htm (visited on 09/16/2015).

Sugden, Robert (1986). *The economics of rights, cooperation and welfare*. Basingstoke Hampshire: Palgrave Macmillan.

Sun, Ron and Isaac Naveh (2009). "Cognitive simulation of academic science". In: *2009 international joint conference on neural networks*. IEEE, pp. 3011–3017.

Swan, T. W. (1956). "Economic growth and capital accumulation". In: *Economic Record* 32.2, pp. 334–361.

Tarko, Vlad (2015). "Polycentric structure and informal norms: Competition and coordination within the scientific community". In: *Innovation: The European Journal of Social Science Research* 28.1, pp. 63–80.

Taylor, Michael (1987). *The possibility of cooperation*. Cambridge: Cambridge University Press.

Tesfatsion, Leigh (2006). "Agent-based computational economics: A constructive approach to economic theory". In: *Handbook of computational economics*. Ed. by Leigh Tesfatsion and Kenneth Judd. Vol. 2, pp. 831–880.

Tesfatsion, Leigh and Kenneth L. Judd (2006). *Handbook of computational economics: agent-based computational economics*. Vol. 2. Elsevier.

Thelwall, Mike and Kayvan Kousha (2016). "Figshare: A universal repository for academic resource sharing?" In: *Online Information Review* 40.3, pp. 333–346.

Thoma, Johanna (2015). "The epistemic division of labor revisited". In: *Philosophy of Science* 82.3, pp. 454–472.

Thomke, Stefan and Eric Von Hippel (2002). "Customers as innovators: A new way to create value". In: *Harvard Business Review* 80.4, pp. 74–81.

Tversky, Amos and Daniel Kahneman (1992). "Advances in prospect theory: Cumulative representation of uncertainty". In: *Journal of Risk and Uncertainty* 5.4, pp. 297–323.

Van Noorden, Richard (2015). "Political science's problem with research ethics". In: *Nature News*.

Van Tuyl, Steven and Amanda L. Whitmire (2016). "Water, water, everywhere: Defining and assessing data sharing in academia". In: *PLoS One* 11.2, e0147942.

Vasilevsky, Nicole A. et al. (2017). "Reproducible and reusable research: Are journal data sharing policies meeting the mark?" In: *PeerJ* 5, e3208.

Veblen, Thorstein (1898). "Why is economics not an evolutionary science?" In: *The Quarterly Journal of Economics* 12.4, pp. 373–397.

— (1899). *The theory of the leisure class: An economic study in the evolution of institutions*. New York: Macmillan.

— (1904). *The theory of business enterprise*. New York: C. Scribner's sons.

Vision, Todd J. (2010). "Open data and the social contract of scientific publishing". In: *BioScience* 60.5, pp. 330–331.
Von Hippel, Eric (1986). "Lead users: A source of novel product concepts". In: *Management Science* 32.7, pp. 791–805.
— (2001). "User toolkits for innovation". In: *Product Innovation Management* 18.4, pp. 247–257.
Von Hippel, Eric and Georg Von Krogh (2003). "Open source software and the "private-collective" innovation model: Issues for organization science". In: *Organization Science* 14.2, pp. 209–223.
Wagner, J. A. (2010). "Open-minded to open innovation and precompetitive collaboration". In: *Clinincal Pharmacology & Therapeutics* 87.5, pp. 511–515.
Walker, Saint John (2014). *Big data: A revolution that will transform how we live, work, and think*. London: John Murray.
Watson, Mick (2015). "When will 'open science' become simply 'science'?" In: *Genome biology* 16.1, p. 101.
Weisberg, Michael and Ryan Muldoon (2009). "Epistemic landscapes and the division of cognitive labor". In: *Philosophy of Science* 76.2, pp. 225–252.
Wenger, Etienne (1998). *Communities of practice: Learning, meaning and identity*. Cambridge: Cambridge University Press.
Wheat, Rachel E. et al. (2013). "Raising money for scientific research through crowdfunding". In: *Trends in Ecology & Evolution* 28.2, pp. 71–72.
White, Hollie et al. (2008). "The Dryad Data Repository: A Singapore Framework metadata architecture in a DSpace environment." In: *Dublin Core Conference*, pp. 157–162.
Wible, James R. (1998). *The economics of science: Methodology and epistemology as if economics really mattered*. London: Routledge.
Wicherts, Jelte M. et al. (2006). "The poor availability of psychological research data for reanalysis." In: *American Psychologist* 61.7, p. 726.
Wilensky, Uri and William Rand (2015). *An introduction to agent-based modeling: Modeling natural, social, and engineered complex systems with NetLogo*. MIT Press.
Wiley, Christie (2018). "Data sharing and Engineering faculty: An analysis of selected publications". In: *Science & Technology Libraries* 37.4, pp. 409–419.
Wilkinson, Mark D. et al. (2016). "The FAIR Guiding Principles for scientific data management and stewardship". In: *Scientific Data* 3, p. 160018.

Williams, Antony J. (2008). "Internet-based tools for communication and collaboration in chemistry". In: *Drug Discovery Today* 13.11, pp. 502–506.

Williamson, Oliver E. (1975). *Markets and hierarchies: Analysis and antitrust implications*. New York: Free Press.

— (1979). "Transaction-cost economics: The governance of contractual relations". In: *The Journal of Law and Economics* 22.2, pp. 233–261.

— (1985). *The economic intstitutions of capitalism*. New York: Simon and Schuster.

— (2000). "The new institutional economics: Taking stock, looking ahead". In: *Journal of Economic Literature* 38.3, pp. 595–613.

Willinsky, John (2005). "The unacknowledged convergence of open source, open access, and open science". In: *First Monday* 10.8.

Wittgenstein, Ludwig (1958). *Philosophical investigations*. Trans. by G. E. M Anscombe. Oxford: Basil Blackwell.

Woelfle, Michael, Piero Olliaro, and Matthew Todd (2011). "Open science is a research accelerator". In: *Nature Chemistry* 3.10, p. 745.

Wohlsen, Marcus (2011). *Biopunk: Solving biotech's biggest problems in kitchens and garages*. New York: Penguin.

Zammito, John (2004). *A nice derangement of epistemes: Post-positivism in the study of science from Quine to Latour*. Chicago, Ill; London: University of Chicago Press.

Zenodo (2019). *About Zenodo*. URL: `https://about.zenodo.org/`.

Zhu, Yimei (2019). "Open-access policy and data-sharing practice in UK academia". In: *Journal of Information Science*, pp. 1–12.

Zollman, Kevin JS (2007). "The communication structure of epistemic communities". In: *Philosophy of Science* 74.5, pp. 574–587.

— (2009). "Optimal publishing strategies". In: *Episteme* 6.2, pp. 185–199.

— (2010). "The epistemic benefit of transient diversity". In: *Erkenntnis* 72.1, p. 17.