The Balancing Role of Evaluation Mechanisms

Cases of Publicly Funded Research Institutions:
MPG, HGF, and CAS

Junwen Luo

The Balancing Role of Evaluation Mechanisms

Cases of Publicly Funded Research Institutions:

MPG, HGF, and CAS

Junwen Luo

Promotion Committee:

Chair:	Prof.dr. T.A.J. Toonen, University of Twente
Secretary:	Prof.dr. T.A.J. Toonen, University of Twente
Promoter:	Prof.dr. S. Kuhlmann, University of Twente
Co-promoter:	Prof.dr. E. Arnold, University of Twente
Assistant promoter:	Dr. H.G. Ordóñez-Matamoros, University of Twente
Members:	Prof.dr. R. Hoppe, University of Twente
	Prof.dr. R. Torenvlied, University of Twente
	Prof.dr. B.J.R. van der Meulen, Rathenau Instituut
	Prof.dr. L. Leisyte, Technical University of Dortmund
	Prof.dr. R. Mu, the Chinese Academy of Sciences

The funding for this PhD project was provided by the China Scholarship Council (CSC).

This thesis was printed with financial support from the Department of Science, Technology and Policy Studies (STePS) of the University of Twente.

Cover design and layout: Liang Ye & Junwen Luo

Cover image downloaded from: <http://hamidslimi.com/wp-content/uploads/2016/02/balance.jpg>

Printed by: Gildeprint - The Netherlands

ISBN: 978-90-365-4121-3

DOI: 10.3990/1.9789036541213

URL: <http://dx.doi.org/10.3990/1.9789036541213>

All rights reserved. No part of this book may be reproduced or transmitted, in any form or by any mean, electronic or mechanical, including photocopying, microfilming, and recording, or by any information storage or retrieval system, without the prior written permission of the author.

THE BALANCING ROLE OF EVALUATION MECHANISMS

CASES OF PUBLICLY FUNDED RESEARCH INSTITUTIONS:

MPG, HGF, AND CAS

DISSERTATION

to obtain
the degree of doctor at the University of Twente,
on the authority of the rector magnificus,
Prof.dr. H. Brinksma,
on account of the decision of the graduation committee,
to be publicly defended
on Thursday 23rd of June 2016 at 14:45 hrs

by

Junwen Luo

born on 20 March, 1987
at Hefeng, China

This dissertation has been approved by the promotor:

Prof.dr. S. Kuhlmann

Prof.dr. E. Arnold

Dr. H.G. Ordóñez-Matamoros

Contents

List of Tables.....	V
List of Figures.....	VI
Acronyms.....	VII
Acknowledgements.....	IX
1 Introduction	1
1.1 What is Organisational Governance of PRIs?	3
1.2 How is Evaluation Used as a Governance Instrument	4
1.3 Why Focus on Evaluation Mechanisms?	5
1.4 Scope of the Study.....	5
1.5 Thesis structure	6
2 Towards an Analytical Instrument—a Basic Conceptual Framework	9
2.1 Organisational Governance	9
2.1.1 The Essence of Organisational Governance for PRIs.....	10
2.1.2 The Stakeholders of PRIs.....	13
2.2 Governance Tensions	14
2.2.1 Tensions within the National Public R&D Landscape.....	14
2.2.2 Tensions within Research Institutions	18
2.3 Evaluation Mechanisms as Governance Instruments	20
2.3.1 The Systemic Nature of Evaluation Mechanisms	21
2.3.2 Institutionalisation of Evaluation Mechanisms	25
2.3.3 Organisational Use of Evaluation Mechanisms as a Balancing Medium	28
2.4 Evaluation Mechanisms Facilitating Organisational Learning.....	32
2.4.1 Knowledge Bases Used by Organisational Governance	33
2.4.2 Organisational Learning Facilitated by Learning Evaluation	35

2.5	Research Question, Sub-questions and Propositions	38
2.6	A Basic Framework on the Role of Evaluation Mechanisms in Organisational Governance.....	39
3	Research Approach	43
3.1	Applying the Three-level Structure into the Case Studies	43
3.2	Operationalisation of the Key Concepts within the Case Studies	46
3.3	Interview Guidelines and Data Collection	49
3.4	Analyses Overview	52
3.5	Validity of the Research Approach	55
4	National and Institutional Contexts of the Three Cases—Towards the Enriched Conceptual Framework.....	57
4.1	National Public R&D Landscapes of Germany and China.....	57
4.2	Institutional Contexts of the Three Cases	63
4.3	Organisational Governance of the Three Institutions.....	67
4.3.1	Strategy	68
4.3.2	Funding	71
4.3.3	Operation	73
4.4	Categories of Stakeholders of MPG, HGF, and CAS.....	75
4.5	The Enriched Conceptual Framework on the Role of Evaluation Mechanisms in Organisational Governance	78
5	Governance Tensions at MPG, HGF, and CAS	85
5.1	Tensions at L1 between Internal and External Stakeholders	86
5.2	Tensions between Headquarters at L1 and Institutes at L2.....	90
5.3	Tensions among Institutes at L2.....	93
5.4	Tensions between Institutes at L2 and Scientists at L3.....	95
5.5	Governance Tensions on Strategy, Funding, and Operation	98

6	Evaluation Mechanisms Respond to and Facilitate Organisational Governance at MPG, HGF, and CAS	105
6.1	Evaluation Mechanisms at MPG, HGF, and CAS.....	105
6.1.1	Evaluation Mechanism at MPG	106
6.1.2	Evaluation Mechanism at HGF	108
6.1.3	Evaluation Mechanism at CAS.....	111
6.2	The Evaluation Mechanisms Responding to Governance Requirements	115
6.3	The Balancing Processes of the Evaluation Mechanisms	120
6.3.1	Institutionalisation of Evaluation Mechanisms at MPG, HGF, and CAS	120
6.3.2	Evaluation Facilitating Stakeholders' Participation and Interaction	129
6.4	Evaluation Mechanisms Facilitating Organisational Learning....	139
6.4.1	Knowledge Bases for Governance.....	139
6.4.2	Organisational Learning Fostered by Evaluation Mechanisms' Learning	141
6.5	The Balanced Organisational Governance	143
6.5.1	Governance Equilibrium Maintenance at L1.....	145
6.5.2	Autonomous Development of Institutes at L2	147
6.5.3	Protection of Freedom and Trust for Scientists at L3	148
7	Conclusions and Recommendations with the Completed Conceptual Framework	151
7.1	Key Results.....	152
7.1.1	What are the Typical Governance Tensions?	152
7.1.2	How are the Governance Conflicts Reflected and Balanced in Evaluation Mechanisms?	154
7.1.3	What are the Balanced Results?	156
7.1.4	What is the Generalisable Principle?	157
7.1.5	The Completed Conceptual Framework	159

7.2	Theoretical Contributions.....	162
7.3	Recommendations	165
7.3.1	Implications for PRI Practices.....	165
7.3.2	Reflections on Public Policy of Research Evaluation	167
7.4	Limitations of This Research.....	170
7.5	Future Research	171
	Appendices	173
	A MPG	173
	A1 Historical dynamics	173
	A2 Organisational structure	174
	A3 Strategy, funding and operation.....	176
	A4 Evaluation practices	178
	B HGF.....	181
	B1 Historical dynamics	181
	B2 Organisational structure	183
	B3 Strategy, funding and operation	184
	B4 Evaluation practices	185
	C CAS	187
	C1 Historical dynamics	187
	C2 Organisational mission and structure	188
	C3 Reforms.....	191
	C4 Evaluation practices	194
	References.....	201
	Summary	217
	Samenvatting.....	219

List of Tables

Table 3.1 Actual subjects visited and interviewed.....	43
Table 3.2 General statistics of the three research institutions.....	44
Table 3.3 Governance characteristics of the three institutions.....	46
Table 3.4 Stakeholders of the three institutions and their concerns.....	48
Table 3.5 Evaluation activities composing evaluation mechanisms at the three PRIs	49
Table 3.6 Interview questions for L1.....	50
Table 3.7 Interview questions for L2.....	51
Table 3.8 Interview questions for L3.....	51
Table 3.9 Overview of interviews conducted.....	52
Table 3.10 Governance characteristics of the three cases	53
Table 3.11 Governance tensions at each level of the three cases.....	53
Table 3.12 Roles of evaluation in governance at each level of the three cases.....	54
Table 4.1 Dimensions of organisational governance in the three PRIs.....	68
Table 4.2 The categories of stakeholders of the three PRIs.....	77
Table 5.1 A summary of governance conflicts at MPG, HGF, and CAS	86
Table 6.1 Characteristics of the evaluation mechanisms of the three PRIs.....	116
Table 6.2 Dynamics of the institutionalisation of the evaluation mechanisms	123
Table 6.3 The balanced results of governance of the three PRIs.....	144
Table B3 The top-down approach of a programme development.....	184
Table B4 Evaluation criteria for the entire programme and each topic in programmes	186
Table C4 The three key evaluation types of CAS before 2012	195

List of Figures

Figure 2.1 A basic framework on the role of evaluation mechanisms in organisational governance	40
Figure 4.1 Institutional dynamics in the German research system between 1990 and 2002.....	59
Figure 4.2 Stakeholders' involvement and interaction concerning governance conflicts	78
Figure 4.3 The enriched conceptual framework on the role of evaluation mechanism in organisational governance.....	82
Figure 5.1 Key governance tensions with two conflicting situations	102
Figure 6.1 An overview of the MPG's evaluation mechanism	106
Figure 6.2 An overview of HGF's evaluation mechanism.....	109
Figure 6.3 An overview of CAS's evaluation mechanism	112
Figure 6.4 Characteristics of the three evaluation mechanisms responding to the governance requirements	117
Figure 6.5 A general process of the institutionalisation of the evaluation mechanisms.....	121
Figure 6.6 Trends of institutionalisation of evaluation mechanisms in the three cases	122
Figure 6.7 Interaction model of MPG's evaluation mechanism.....	133
Figure 6.8 Interaction model of HGF's evaluation mechanism.....	134
Figure 6.9 Interaction model of CAS's evaluation mechanism	137
Figure 6.10 Evaluation mechanisms facilitate organisational learning.....	143
Figure 7.1 The completed conceptual framework on the role of evaluation mechanisms in organisational governance	162
Figure 7.2 Reflection on public policy of research evaluation	168
Figure A2 Organisational structure of MPG	175
Figure B2 Organisational structure of HGF	183
Figure C2-1 China's R&D landscape	189
Figure C2-2 Organisational structure of CAS.....	190
Figure C4 The three-level evaluation indicator system of CAS	197

Acronyms

AvH: Alexander von Humboldt Foundation
BMBF: Federal Ministry of Education and Research
CAS: Chinese Academy of Sciences
CASS: Chinese Academy of Social Sciences
DAAD: German Academic Exchange Service
DFG: German Research Foundation
EFI: Commission of Experts for Research and Innovation
EU: European Union
FhG: Fraunhofer Society (Fraunhofer–Gesellschaft)
GDP: Gross Domestic Production
GERD: Gross Domestic Expenditure on R&D
HGF: Helmholtz Association (Helmholtz–Gemeinschaft Deutscher Forschungszentren)
HR: Human resources
HRK: German Rectors' Conference (Hochschulrektorenkonferenz)
KIP: Knowledge and Innovation Programme
MLP: Medium– and Long-term Plan
MOST: Ministry of Science and Technology
MPG: Max Planck Society (Max–Planck–Gesellschaft)
MPI: Max Planck Institute
NIS: National Innovation System
NPM: New Public Management
NSFC: National Natural Science Foundation of China
OECD: Organisation for Economic Co-operation and Development
PAP: Pioneer Action Plan
POF: Programme-oriented funding
PRI: Publicly Funded Research Institution
R&D: Research and Development
SAB: Scientific Advisory Board
S&T: Science and Technology
WGL: Gottfried Wilhelm Leibniz Science Association
WR: Council of Science and Humanities (Wissenschaftsrat)

Acknowledgements

Finally I move to the step of writing acknowledgments which is, on the one hand, the easiest part to write compared with the other chapters. On the other hand, the two-page acknowledgements present my most important findings in the past four and half years' PhD journey, far beyond my research results.

First of all, my research capabilities especially communication and writing skills cannot grow so fast without the practice with and kind help from my supervisors and colleagues. My promoter Stefan, co-promoter Eric and daily supervisor Gonzalo always guide my way patiently in a very professional and critical manner, especially in the darkest writing steps of this journey. My resourceful colleagues at STePS, Lissa, Peter, Alejandro, Carla, Haico, Katrin, Bart, Maria, Pierre, Gustavo and many others, discussed my research as well as other interesting topics, which gave me useful feedback and rich experiences. Our discussions make me more and more confident as an independent researcher! Also, our department secretary Evelien, Hilde, and Marjatta, gave me lots of help in many administration activities. Each of you—my knowledgeable and lovely colleagues—taught me kinds of wisdom you may not notice yourselves.

Great thanks to the China Scholarship Council (CSC) which funded my PhD project, as well as the Netherlands Organisation for Scientific Research (NWO) which funded my first trip from China to Netherlands. That 'Talent Training China Match-making Interview' was my first trip going abroad which realised many match-makings: my academic relationship with University of Twente, my friendship with several other candidates, particularly, my romantic relationship with one of them who is my husband now! We may not have the chance to meet each other without this trip. In addition, my British dream came true thanks to the EU-SPRI forum which funded my three-month PhD circulation in Manchester Institute of Innovation Research of Manchester Business School. Those colleagues in Manchester: Jacob, Phil, Abdulla, Yanchao, Li Chao, Xiuqin, Siobhan etc. helped me a lot during my stay. Thanks to my department—STePS—funded my several international conferences and research field works in both Germany and China.

I appreciate the availability and generosity of the people I interviewed in the three big and famous research institutions I studied: Max Planck Society (MPG) and Helmholtz Association (HGF), both in Germany, and the Chinese Academy of Sciences (CAS). Without your contribution of interview data and feedback on my

thesis, I was not able to conduct such a qualitative empirical research to open some black boxes to other scholars. The interviewees' names are not included in this thesis but kept well in my mind.

I luckily have two professional proof readers, Jillian Yeow from UK, and Zethyn Ruby from US, thank you both very much for your careful language edition and correction. Otherwise, my Chinese-style English can never satisfy my supervisors and committee members. Of course I take full responsibility of every mistake or imperfection still present in the text.

I am grateful to my dear Chinese friends around in Netherlands. We spent so many colourful and amusing moments together during my PhD journey: travelling around Europe, making dinner parties on weekends, shopping, camping and so on. Wenlong, Xuemei, Meiru, Sun, Jiahui & Liu Yao, Du Ying & Lantian, Mengdi & Youwen, Xiaohua & Kenan, Aijie & Wang Yi, Xichen & Yujie, and many others, you may not know that at our gathering time, I often observed you—young and promising scientists—like other observant in my research. You guys always surprise me by your intelligence and sense of humour. I wish all of you the best in your scientific career and personal lives.

My beloved and dearest husband, Liang Ye, you bring me so much love, peace, patience, hope, confidence, wisdom, knowledge, tolerance, and enthusiasm. We shared both the happiest and the most difficult moments in this journey hand by hand. I hope we will always do so in our future — love, admire, and trust each other.

I always feel lucky to have such open-minded parents, Luo Zhengcai and Peng Xiaohong, thanks for your unconditional love, trust and support which are always my strongest backup force no matter what happens. Our one-month trip in Europe was as beautiful as a dream which helped me relax so much after exhausting writing. I know you will always be there. I love you!

感谢我的父母，罗正才和彭小红，对我无条件的爱和支持，你们永远是最坚实的后盾！我爱你们！

1 Introduction

‘When making decisions about funding, hiring, tenure, or promotion, make assessments based on scientific content rather than publication metrics’.

---- **DORA (Declaration on Research Assessment, 2012)**¹

How can research be evaluated? This is a complex question involving diverse concerns. As DORA shows, the current publication metrics, for example, focus on research output but devalue scientific content. This leads to inappropriate decision-making on funding, hiring, tenure, promotion, etc., which are harmful to the development of scientific research at various levels. There are concerns regarding research assessment on at least two levels—at the national level, the wide research and development (R&D) landscape, and the organisational level at research institutions. At the national level, it is difficult to implement an overarching research evaluation system with standardised assessment criteria that is suitable for all disciplines (like physics or chemistry), fields (like health or environment), or for research institutions that are different in kind or size. At the organisational level, the challenge is different. Indeed, there are ways to conduct research evaluation that follow a set of principles and practices defined centrally and implemented in a distributed way. Arguments in supporting that view abound in the management literature referring to, for example, transaction cost, efficiency, and organisational governance. Based on this, a series of questions could be proposed: Can there be standardised research evaluation within publicly funded research institutions (PRIs)? Who will do such work? And what value would be produced?

At the institutional level, a typical crucial question within PRIs is how to allocate public resources efficiently to produce the best research for society, which involves, inter alia, measuring scientific content and output of the various working units under the organisational umbrella. According to Suttmeier et al. (2006, p. 58), inappropriate resource allocation in science includes a shortage of institutional block funding, low salaries for scientists, non-research personnel overstaffing, ageing researchers with limited potential and outdated research priorities. In the attempts to overcome these problems, tensions emerge among various stakeholders. Who is going to solve those tensions and how? Can they be avoided

¹ DORA, initiated by the American Society for Cell Biology (ASCB) together with a group of editors and publishers of scholarly journals, recognises the need to improve the ways in which the outputs of scientific research are evaluated (Way & Ahmad, 2013).

or anticipated? This is not only a problem for organisational leaders, but there is also a lack of knowledge in the current literature on ways to address such concerns in the practices of PRIs. Therefore, why not look at the experiences of leading PRIs worldwide that operate in strong economies and embrace complex organisational characteristics? What can be learnt from them?

Many national PRIs establish their own organisational governance schemes that reflect the changing governance of the sciences (Whitley & Gläser, 2007). The National Institutes of Health (NIH) in the US, the Spanish National Research Council (CSIC), the French National Centre for Scientific Research (CNRS), the German Max Planck Society (MPG), the German Helmholtz Association (HGF), and the Chinese Academy of Sciences (CAS) are examples of PRIs. This study selected MPG, HGF, and CAS as subjects because of their similarities and comparability, as well as the capability and availability of the author to access to their unpublished information. They are the world's leading institutions for high-quality research and cover the diverse research orientations defined in this thesis—fundamental, strategic and applied research. According to the *Nature Publishing Index* (2013)² which measured research performance by high-quality publication, CAS ranked first among international research institutions, MPG ranked third, and HGF eighth. Does this mean CAS performs better than MPG and HGF? Can we read performance of all PRIs on organisational governance behind the publication numbers?

To contribute evidence and reflection to partly answer the above general and complex questions, this thesis explores the role of evaluation mechanisms in organisational governance of PRIs through case studies of MPG, HGF, and CAS. Of course, the three PRIs have quite different characteristics of organisational governance so that their other differentiated evaluation practices beyond publication are supposedly diversely implemented. It is worthwhile to study whether there is a common principle behind their varying evaluation practices and what kind of role their evaluations play in their different organisational governance.

Sections 1.1 to 1.3 will clarify briefly the key concepts of this study and their mutual dependencies, including PRIs, organisational governance and evaluation

² The *Nature Publishing Index* is an international ranking of top level research institutions based on the number of articles published in *Nature's* world renowned research journals. See http://www.nature.com/nature/journal/v515/n7526_supp/fig_tab/515S98a_T2.html for the Top 200 institutions in 2013.

mechanisms. A basic set of questions and propositions will be included in the scope of the study in section 1.4. Section 1.5 will show the structure of the thesis, that is, the work presented in each chapter.

1.1 What is Organisational Governance of PRIs?

First, the concept of *institution* when referring to PRIs includes both organisations as players and rules (procedures, arrangement, routines, habits, and norms) to be followed by the organisations (North, 1991). Therefore, in this thesis, PRIs refer to concrete research organisations with their own rules for doing research and development (R&D) activities. To distinguish *public* research institutions from other research institutions, the former have to respond to external requests for justification of public investment to obtain public support because of their public nature.

According to Jansen (2007, p. 233), governance within research organisations can be understood from multiple perspectives of political science, law, economy and sociology. As a brief introduction, organisational governance of PRIs in this thesis relates to *the processes whereby organisational objectives are set and pursued, and policies and decisions on strategy, funding and operation are made and monitored*. The detailed definition will be discussed in section 2.1.1. In particular, organisational governance of PRIs is embedded in the national public R&D landscape and in three levels of institutional structures (macro-, meso- and micro-levels), all embracing multiple stakeholders. Understandably, organisational governance of PRIs can be very complex and involve hierarchical layers, institutes at multiple locations and working on various research disciplines and fields and organisational cultures, etc. Indeed, given such complexity, the organisational governance of PRIs is a big challenge.

Characteristics of the organisational governance of PRIs are still black boxes for the external researchers like us. For instance, how do they cope with scientific risks in various fields that are working towards the overarching strategies? Is there funding competition between the fields and institutes? Which operation processes are in place that the heterogeneous research units and facilities face? The answers could be complicated for various stakeholders—internal institutes and scientists, external funding bodies, the scientific community, collaborating universities and enterprises, and the general public as taxpayers. The stakeholders are concerned with different

governance aspects with different powers. Therefore, tensions arise and have driven institutional dynamics over years.

1.2 How is Evaluation Used as a Governance Instrument

Evaluation has become a professional practice, with clear achievements and competencies that can be taught (e.g. Guba & Lincoln, 1989; Kuhlmann, 2015). Evaluation is becoming integrated as an administrative routine at many levels and in many types of institutions (Hellstern, 1986). Research institutions perform evaluations in order to demonstrate their trustworthiness to the outside world (Hasson, 2006, p.159). However, the question of whether evaluations advance organisational governance has been overlooked (Simon, 2013). Some studies have recognised the lack of systemic consideration of institutional contexts in evaluations (Edler, 2012; Højlund, 2014; Raina, 2003). Hasson (2006, p. 169) emphasises the necessity to understand research activities and their evaluation in the context of research organisations and to make the governance question visible via evaluation. The focus, when explaining the role of evaluation mechanisms, needs to be on the evaluating organisation and its conditioning factors, rather than on the evaluation itself (Højlund, 2014).

In the wider context of science policy-making, science policy-makers request evidence about the effectiveness and impacts of public R&D policies, programmes and improved approaches to evaluate them (Feller, 2007). Performance-based research funding systems have been institutionalised in many countries like the UK and Australia (Gläser & Laudel, 2007, p. 253), but not in Germany (Jansen, 2007, p. 234) or China (Grayson, 2015, p. 5). Within performance-based research funding systems, the use of only quantitative evaluation focusing on publications and patents to determine the amount of institutional funding of PRIs has been criticised (Good et al., 2015, p. 100). One of the reasons is that an important factor of public research—societal impact—is much harder to measure and forecast in basic research than in applied research (Cozzens, 2007).

Although globally there are interests in and critics of different types of evaluation, both internal and external drivers for evaluation of PRIs always concentrate on research quality and the impact of related resource allocation (Cicchetti, 1991; Gläser & Laudel, 2007; Hartmann & Neidhardt, 1990). Peer reviewing is widely accepted not only because of professional standards, but also because of 'a lack of better alternatives to evaluate scientific research' (Simon & Knie, 2013, p 406). For

PRIs, evaluation is no longer used mainly to judge that research is working well or badly, but also as an analytical framework relevant for decisions regarding future-oriented strategies, reward structures, personnel policy, etc. (Feller, 2007). A variety of dimensions of research evaluation are increasingly discussed by literature for better governance of publicly funded research: periodicity and cost, the unit-of-result analysis, performance and output measures that consider disciplinary differences, and percentages of funding allocated according to categorised evaluation results (Cozzens, 2007). Consequently, for the long-term development of PRIs, evaluation is increasingly used as an instrument to respond to governance requirements.

1.3 Why Focus on Evaluation Mechanisms?

The rationale for evaluations impacting organisational governance leads us to the question: What are '*evaluation mechanisms*'? Evaluation mechanisms in this study are defined as *the ways the whole evaluation system is constructed by all individual evaluation activities and operate vis-à-vis the complex institutional environment and stakeholders*. An overarching mechanism is supposed to self-improve by individual evaluation activities supplementing each other (Mark et al., 1999; Vega & Chiasson, 2015). For example, functions of qualitative research evaluation for scientific recommendations and of quantitative performance measurement of research institutes for allocating institutional funding (Tsay et al., 2003) are combined. Evaluation mechanisms could work at both national and institutional levels for particular governance purposes. At institutional level, evaluation mechanisms of individual PRIs are expected to be used as governance instruments by providing reliable inputs for decision-making and by facilitating organisational learning, which will be explained in the next chapter.

1.4 Scope of the Study

This study originated from my personal interest in R&D evaluation in general and particularly in Germany and China. The inspiration came from witnessing a lack of literature on the organisational use of evaluation for PRIs' governance. Governance conflicts under national systems and institutional contexts were found by this study to be underestimated. The national R&D landscape of Germany has well-organised and clearly delineated PRIs and that is what China's R&D landscape needs to learn from. Three case studies—MPG and HGF in Germany and CAS in China—are conducted. Their comparative analyses show the different dynamics of organisational governance and evaluation mechanisms. Based on that, this study

aims to increase the understanding of the role of evaluation mechanisms as governance instruments of PRIs in general and beyond the three cases.

An overarching and complex research question is proposed: ***How do evaluation mechanisms of PRIs play a balancing role and help to mitigate tensions in organisational governance?*** To answer that, a conceptual framework will be built in an abductive way for better understanding the relationships between organisational governance and evaluation mechanisms. Hypothetically, there are bridges between these two concepts. The first is that governance tensions justify and influence evaluation mechanisms; the second is that evaluation mechanisms reflect and balance governance conflicts. Studying these double linkages is necessary not only for better understanding the governance and evaluation of PRIs as discussed in numerous but disconnected literature, but also for providing stepping stones for theory—building on how to balance stakeholders' tensions and facilitate organisational learning.

1.5 Thesis structure

The way this thesis is organised is also the way the conceptual framework is developed to understand the co-evolution of organisational governance and evaluation mechanisms of PRIs. The process to establish such a conceptual framework goes from a basic framework to an enriched one, then to the final comprehensive one.

Chapter 2 interprets the literature related to the key concepts of organisational governance, evaluation mechanisms and their relationship, and leads towards an analytical instrument in this thesis. It is proposed that the case studies test the four sub-questions deriving from the overall research question and the corresponding propositions. Based on that, a basic conceptual framework is established to guide the following empirical studies.

Chapter 3 outlines the research approach and explains the operationalisation of the concepts discussed in Chapter 2. Methods of collecting and analysing interview data are explained. The validity of this study is discussed.

Chapter 4 presents the national and institutional contexts of MPG, HGF, and CAS—their challenges related to the R&D landscapes in Germany and China respectively, and characteristics of their organisational governance. Furthermore, their shared categories of stakeholders embedded in the national and institutional contexts can

lead to their hypothetical conflicts. Based on that, an enriched conceptual framework is proposed with updated sub-questions and propositions to be tested in Chapters 5 and 6.

Chapter 5 presents the empirical findings on the categorised governance tensions of the three PRIs, both activities-wise (strategy, funding, operation) and level-wise (L1, L2, L3). Those conflicts are closely related to the diverging interests and power of the multi-level stakeholders. Interactions between the stakeholders across the different levels of governance (L1, L2, L3) in each PRI are highlighted. Some tensions are shared by the PRIs, and some are unique.

Chapter 6 investigates comparatively the balancing processes and results of the three evaluation mechanisms implemented at MPG, HGF, and CAS. Although the three evaluation mechanisms perform and are institutionalised differently, they all respond to the governance tensions and facilitate organisational learning in their own contexts. The balanced governance results for the three cases are found to have similar principles at each organisational level (L1, L2, L3). These findings from the case studies validate and also contribute to complete the analytical framework in the next chapter.

Chapter 7 gives conclusions and recommendations based on the above empirical findings. A comprehensive analytical framework is completed with all the elements and their relationships that are found. The key empirical results on MPG, HGF, and CAS are organised to answer the four sub-questions proposed with the enriched conceptual framework in Chapter 4. A generalisable principle on the balancing roles of evaluation mechanisms in organisational governance beyond the three cases is found—the contextualisation of designs and uses of evaluation mechanisms for reflection and mitigation of governance tensions. Then the theoretical contribution of this study to current literature is presented. Therefore, both explicit implication about organisational practices of governance and evaluation mechanisms, and implicit reflections about public policy-making in the wider public R&D landscape are discussed. The limitations of this study and some ideas for future research are presented at the end.

The appendices at the end of the thesis provide supplementary materials about the three PRIs studied related to their historical dynamics, organisational structure, governance characteristics and evaluation practices.

2 Towards an Analytical Instrument—a Basic Conceptual Framework

What evidence could be found in the current literature on the relationship between *organisational governance* and *evaluation mechanisms*? Is there a satisfactory analytical framework addressing, assessing, and accounting for or explaining such relationship? If not, how can the current understanding and literature help to create a framework to respond to the overarching research question? Section 2.1 discusses the literature about *organisational governance* regarding its definition for publicly funded research institutions (PRIs) and the stakeholders involved in PRIs. Section 2.2 focuses on *governance tensions* in national and institutional contexts of PRIs. At both national and institutional levels, the systemic nature (section 2.3.1) and the institutionalisation process (section 2.3.2) of *evaluation mechanisms* are discussed in relation to governance requirements in general and at PRIs in particular. Then literature on organisational use of evaluation as a governance instrument is discussed in section 2.3.3. Section 2.4 addresses how evaluation mechanisms can facilitate organisational learning. Based on these theoretical discussions, further specified research questions and propositions are presented in section 2.5.

Consequently, after realising the lack of a proper analytical instrument to satisfactorily account for the relationship studies, a basic conceptual framework for answering the overarching research question is sketched in section 2.6. This basic framework will be validated and enriched in the subsequent empirical chapters 4, 5, and 6, using the cases of two German PRIs and one PRI from China. It should be noted that not all the conceptual discussion in this chapter contributes explicitly to the basic framework, as some conceptual elements will be added to the following frameworks if they are validated in the empirical chapters. This is a consequence of the abductive approach followed to answer to the research questions and address the propositions suggested in this thesis.

2.1 Organisational Governance

‘The term “governance” is a very versatile one’—as the European Commission illustrates the problem of defining it (Jansen, 2007, p. 103). Indeed, as introduced above, the disciplinary concept of *organisational governance* can be understood from multiple perspectives and can be complex in practice. This thesis focuses on its essence at PRIs and its related challenges and tensions.

2.1.1 The Essence of Organisational Governance for PRIs

According to Jansen (2007, p. 12), *governance* in research organisations is central to analysing both the ongoing reforms in a research system involving universities and non-university PRIs, and the effects on research behaviour and decisions of research organisations, groups, and individual researchers. However, a useful and adequate theory of governance of research organisations applicable to both research systems and organisations is lacking (Jansen, 2007, p. 14). This thesis explores the organisational governance of PRIs embracing two basic dimensions—institutional characteristics of national research systems and internal research orientations, rules, and decisions at organisations, groups, and by individual researchers. These two dimensions also determine the way in which the following conceptual and empirical analyses of organisational governance at PRIs are presented. This leads us to ask a question: in which specific aspects do the two basic dimensions of *organisational governance* perform?

According to Strong et al. (2008, p. 152), to maximise the benefits and efficiency of research activities, organisational governance of PRIs should specify the *governance-related decisions and the actors* who are expected to make them. Organisational governance is defined by Hermanson and Rittenberry (2003, p. 27) as ‘a process dealing with the procedures utilised by the representatives of the organisation’s stakeholders to provide oversight of risk and control processes administered by management’. This definition, applicable to PRIs, involves the key elements of recognising stakeholders and their activities on monitoring, risk management, and control. As Heinze and Kuhlmann claim (2008, pp. 889–890), organisational governance of PRIs involves distinguishable forms of institutional coordination of autonomous but interdependent units (like research institutes and groups) and stakeholders subject to different types of rules: hierarchy, competition, network, association, and community. These elements and rules guide the exploration of stakeholders’ specific activities on *organisational governance* of PRIs. This thesis focuses on the three aspects of organisational governance: **strategy, funding, and operation**. The choice to categorise governance-related decisions and challenges does not follow any specific literature but is inspired by various governance-related aspects used in organisational and institutional theories (Porter, 1996; Guerrieri & Tylecote, 1997; Peters & van Nispen, 1998). The following is a discussion about these three aspects as found in related literature.

According to Mintzberg (1994), the **strategy** of organisations is expected to bridge the gap between policy and tactics normally in four ways: plan, pattern in actions over time, position, and perspective vision and direction. Based on Porter's work (1996), strategic planning and management of PRIs are supposed to be highlighted by their governance arrangements to achieve their unique and valuable position and long-term development. Some new organisational strategies emerge over time because 'intentions of strategies always collide with and accommodate changing realities' (Mintzberg, 1994, p. 25). The competitive advantage of firms include strategic management quickly 'identifying new opportunities and organising effectively and efficiently to embrace them' (Teece et al., 1997, p. 509). However, many research organisations in much of continental Europe have 'heavy administrative shells by ministries of education and science' (Whitley & Gläser, 2007, p. 10), so they are considered to focus more on short-term strategies to solve immediate problems (Good et al., 2015).

Second, according to Guerrieri and Tylecote (1997, p. 113), a **financial** system with principles of resource accumulation and allocation is considered 'an underlying requirement for governance advantage in successful innovation'. According to their visibility of external financial resources, various kinds of investment of PRIs can be ranked: fixed facilities with the highest visibility, then R&D expenditure, and distribution with the lowest visibility (Tylecote, 1994). In addition, PRIs have internal financial systems with stable relationships between internal governors and external funding bodies. Stability of institutional funding helps to guarantee attraction and retention of high-quality scientists (Good et al., 2015; Grayson, 2015). According to Cruz-Castro and Sanz-Menendez (2007, p. 212), PRIs have historically had less autonomy and greater presence of block funding than universities, but this is changing over time because of the funding competitions and the related evaluation standards.

Furthermore, some countries like the UK and Australia introduced national metrics-based research assessment to determine funding allocation (Geuna & Martin, 2003). For both Germany (Jansen, 2007, p. 234) and China (Fang, 2012, p. 30), which weight research results less in the overall context, how do they decide and allocate research funding to the PRIs? This will be explored in the case studies. For the time being, it may be sufficient to say that the financial reforms in the German research system—introduction of programme-oriented funding at some PRIs—are intended to create a stronger orientation towards performance and externally

evaluated programmes (Jansen, 2007, p. 234). China also has made significant reforms on research funding allocation, for example, through the Knowledge Innovation Programme (KIP) (Suttmeier, 2006, p. 58). However, whether and to what extent the financial reforms have influenced the organisational governance of PRIs is to be explored empirically.

Third, according to Arundel et al. (2007), **operation** refers to the process of making a system work towards its strategies, which involves particular organisational structures, arrangements, and actors; this can also be applied to PRIs. Porter (1996, p. 3) proposes that constant improvement in operational effectiveness is necessary to achieve organisationally superior output (which is profitability in business in his work). This thesis integrates further dimensions discussed in governance literature into the aspect of operation—academic self-organisation, hierarchal self-management, networks, and cultural ordination (Jansen, 2007, p. 109). The broad involvement of *operation* considers the broad definition of governance as the regulation of the coordination of various organisational actions (Benz, 2004). Based on the work of Heinze and Kuhlmann (2008), it is assumed that there are more operational targets shared by all PRIs than strategic and financial concerns, such as collaborating and training activities.

Therefore, *organisational governance* of PRIs in this thesis is defined as ***the processes whereby organisational objectives are set and pursued, the activities and rules of research and researchers are produced, sustained, and regulated among stakeholders in relation to governance arrangements on strategy, funding, and operation.*** Nevertheless, as complex adaptive systems, how organisations can adapt effectively to their environment is widely discussed in the field of strategy and organisation science (Baumann, 2015, p. 169). A dynamic perspective is also required towards the analyses of organisational governance. Are those governance aspects (strategy, funding, operation) stable or changing dynamically? Governance activities evolve dynamically through the invention of new principles and practices (Daft, 2010) and from organisational learning (March, 1991), which can result in collective action and coordination between arenas (Benz, 2007, p. 10) and stakeholders (Bryson et al., 2011). The dynamics of organisational governance will be explored empirically considering specific national and institutional contexts, and particular stakeholders.

2.1.2 The Stakeholders of PRIs

Stakeholders' roles are considered key factors in both literature on *organisational governance* and in the analyses done throughout the thesis. Stakeholder theory, developed greatly by Freeman (1984), has gained wide acceptance since the 1980s in organisation theory and design (Daft, 2010), business strategy (Porter, 1996), organisation management (Bryson et al., 2011), strategic management (Freeman, 1984; Teece et al., 1997), corporate governance (Zhang et al., 2014; Young et al., 2008), and public policy (Hoppe, 2011). Freeman (1984, p 46) defined a stakeholder as 'any group or individual who can affect or is affected by the achievement of the organisation's objectives'. A definition of stakeholders, from public and non-profit sector literature, is provided by Eden and Ackermann as 'people or small groups with the power to respond to, negotiate with, and change the strategic future of the organisation' (Eden & Ackermann, 1998, p. 117).

Considering the definitions of both *organisational governance* and *stakeholders*, stakeholders of PRIs can be categorised as coming from external systems and internal organisations, groups, and researchers. The internal ones include decision-making committees, scientific advisory boards, headquarters, and departments responsible for operation, staff at member institutes, and researchers. External ones include members of government units and ministries—funding bodies, the scientific community, research councils, and collaborators like enterprises and universities, and the general public as taxpayers (Whitley & Gläser, 2007; Biber-Klemm, 2011). In relation to the governance aspects (strategy, funding, operation), each categorised stakeholder group has its own perspective.

As Biber-Klemm (2011) claims, in principle, stakeholders of academic research should take responsibilities as both members of the research community to create trust, transparency of research, and cooperate with partners, and as members of civil society to advocate for monitoring research. Those dual responsibilities seem to reflect an internal conflict for each stakeholder of PRIs. According to Dahler-Larsen (2011, p. 55), the interests of various stakeholders are often controversial because of their utility—maximising of their own interests, which relates to the principle-agent tensions discussed by Braun and Guston (2003). Therefore, governance tensions can be assumed to result fundamentally from the perspectives of stakeholders who cause them—their multiple responsibilities and

contributions, interests, and power. Which stakeholders cause which kind of governance tensions will be discussed in the next section.

2.2 Governance Tensions

According to the categorised external and internal stakeholders of PRIs, governance tensions can also be categorised at national (or system, like the European Union (EU)) and organisational levels, which will be discussed respectively in the following two sub-sections.

2.2.1 Tensions within the National Public R&D Landscape

Since organisational governance of PRIs includes their institutional characteristics of the national research systems, the system-level tensions can influence organisational governance of PRIs—the study subject. Key PRIs are considered by Borrás and Edler (2014, p. 11) as organisational actors of a national innovation system that produce, adopt, diffuse, and use knowledge and technologies. All PRIs within one country contribute to a national public R&D landscape embedded in the wider national innovation system, and are positioned between industrial and educational systems (Kuhlmann & Arnold, 2002). National Innovation Systems (NISs), an important concept in governance studies, are defined as networks of institutions in the public and private sectors, inside the borders of a nation, the activities and interactions of which produce, diffuse and use new technologies (Freeman, 1987; Lundvall, 1992). A successful NIS provides an institutional context with platforms and policies to support the interactions and knowledge flows of its constituting PRIs (Raina, 2003). However, the NIS concept itself is controversial because of its focus on a national system versus the context of globalisation (Lundvall et al., 2002). Such controversy between the focus on research institutions and their international communities of research also exists in the definition of PRI governance.

Due to their *public* nature, PRIs embrace strictly defined missions from public funders within the national R&D landscape. In this sense, PRIs are supposed to deliver excellent research and also are required to provide scientific evidence for national policy-making. The realisation of their mission comes along with an increasing public accountability under the principles of New Public Management (NPM) (e.g. Hood, 1995) where citizens are viewed as customers and public organisations as public managers or agents. In fact, NPM has given rise to rigorous reforms in Western countries and particularly increased competitive pressures

within and among PRIs because of increased scarcity of available financial resources (Schimank, 2005). Since the early 2000s, throughout Europe, research policies and institutions have been deeply influenced by NPM to emphasise efficiency and quality of research products and services (Simon & Knie, 2013).

However, these *standardised missions* sometimes conflict with the *heterogeneous scientific development* of various PRIs with different research orientations—fundamental, strategic,³ and applied ones—considering purposes, approaches, and impacts (Arnold, 2004; Hage et al., 2007; Irvine & Martin, 1984; McCarthy, 2011; Whitley & Gläser, 2007). Such a classification and the related characteristics of research orientations are used for evaluating the relative positions and products of various PRIs in a national R&D landscape. A mission-oriented public research system, for example, uses evaluation as a means for policy analysis with increasing emphasis on societal grand challenges (Amanatidou et al., 2014). However, in basic research, serendipitous results are often the most interesting and ultimately may have the most value and bring about breakthroughs, but it is hard to predict and assess their societal benefits early on (OMB, 2005, p. 60).⁴ The administration of OMB in the USA expects agencies to improve the research management under different research orientations and adopt them according to management practices (Cozzens, 2007). Highly differentiated research systems not only mean that scientists do different kinds of R&D activities that should be evaluated on different criteria, but also that they understand and accept such divisions of work.

National governments always play a major role in supervising PRIs by imposing standards and regulations with strong performance assessment tools (Good et al., 2015) that sometimes do not fit PRIs' *complex missions*. As a result of changes in the wide R&D landscapes from political pressures, the increasingly complex missions of PRIs, over time, require the PRIs' to have stable roles which have been adopted dynamically in given contexts (Bozeman & Crow, 1995). Evaluation of these complex PRIs has become a typical and frequent way in many countries to demonstrate that research investment is effective and delivers public benefits (Boswell, 2008; Meyer & Rowan, 1977), especially those with relatively strong research funding, including Germany and China (Fang & Tian, 2012; Krull, 1995). In

³ Strategic research is defined as, 'basic research carried out with the expectation that it will produce a broad base of knowledge likely to form the background to the solution of recognised current or future practical problems' (Irvine & Martin, 1984, p. 4).

⁴ OMB: the US Office of Management and Budget.

fact, the most powerful and widely used approach is *ex-ante* peer review in publicly funded project and programme evaluations through which research funders steer and shape research (Cicchetti, 1991).

Therefore, a conflict between *policy-based regulatory science* and *curiosity-driven and open-ended academic science* has been found by Strassheim and Kettunen (2014, p. 266). The concept of *regulatory science policy* has been suggested (Jasanoff, 1995) to characterise PRIs' contribution to the target of carefully framing evidence to be politically relevant and justified for use in policy-making. Regulatory science policy has embraced the widespread acceptance in international research on science policy interaction (Johnston, 2012) as a general prescription for accountability and better governance of science policies (Davies et al., 2000; Nutley et al., 2010). However, regulatory science policy involves many external stakeholders (such as government and industry) in science policy interaction and produces policy-driven R&D under complex institutional and statutory deadline constraints; this process has high-level and diverse conflicting interests (Jasanoff, 1990; Schiene and Schimank, 2007). The demarcation between science and politics is thus uncertain for PRIs and their researchers, leading to conflicts between the justifiability of scientific expertise and policy-driven scientific evidence.

Moreover, public legitimisation calling for more transparency of R&D investments of PRIs could go against trust in academic freedom⁵ and scientific freedom.⁶ Public awareness, engagement, and acceptability of science lead political actors to seek new ways to legitimise research investment by robust standards and guidelines; see for example, the Dutch government's recent effort to create legitimisation for public research funding⁷ by inviting academics, industry, civil society organisations, and citizens to submit research questions; a German governmental major participatory foresight effort to cope with society's expectations vis-à-vis publicly funded research in the early 2000s ('Futur', see Smits & Kuhlmann, 2004); a horizontal foresight to address societal challenges in Danish priority-setting for public funding of strategic research two years ago (Nielsen, 2014). Given limited

⁵ Academic freedom (to teach or communicate ideas or facts) and tenure are often attacked for political reasons, especially in the behavioural sciences (Ceci, Williams, & Mueller-Johnson, 2006).

⁶ Scientific freedom usually refers to the idea of freedom applied to natural science, in particular the practices of scientific research and discourse mainly by publication (Smith, 2004).

⁷ <http://www.wetenschapsagenda.nl/?lang=en>.

public resources, PRIs claim to address so-called grand challenges collaboratively. Greater autonomy beyond research performance assessment is always expected by PRIs so that they can develop their profiles and strategic abilities to compete with each other (OECD, 2013). There have been discussions around whether public investment in science should move from pure faith in science, because of its complex and non-transparent nature (Borrás & Edler, 2014), to trust based on the scrutiny of science and scientists that can be understood outside science (Cozzens, 2007). For example, the evaluation methodology in the Czech research system, strongly dependent on quantitative research assessment for public resource allocation, can be seen as a symptom of the lack of trust in effective governance of the national public R&D landscape (Good et al., 2015).

The bolstered public accountability gives rise to *principal-agent problems* as a typical tension in PRI governance. The principal-agent theory shows that a lasting tension in research policy-making is the problem of delegation because of lack of information on the side of non-scientists leading to typical problems of adverse selection and moral hazard of scientists (Braun & Guston, 2003). In the early 1990s, Braun introduced the principal-agent concept in the context of research policy-making, referring to the untrusting relationship under double circumstances. The first is between policy-makers as principals, and the various funding agencies responsible for the implementation of research policy as agents; and the second is between funding agencies as principals and scientists as agents. The balancing actions in-between demand a considerable degree of independence in relation to the principals and to the third party (Rip, 2006; Van der Meulen & Rip, 2000).

In this context, intermediate research councils as third parties were considered to function as '*aggregation machines*' to bridge the gap between ongoing knowledge production and the research agenda of sponsors (Rip, 2000, p. 467). This machine evaluates proposals and decides their funding to balance the often-opposed interests of policy-makers and scientists, and those between risk-averse decisions and the recognition of developmental potentials of the proposals (*ibid*). To avoid the classic situation of free riders in academia, that balance must rely on a learning-based combination of managing and evaluating research in research councils (Hansson, 2006). For instance, the interactive learning (Lundvall et al., 2002) of the aggregation machine which interact with policy-makers and scientists respectively, help to facilitate symmetry of scientific information between research policy-making and scientists.

In a nutshell, some governance tensions in the national public R&D landscape are identified by the literature. Meanwhile, national research evaluation is sometimes analysed as a relevant instrument when coping with system-wide tensions. Whether national-level tensions affect organisational governance of PRIs, and whether evaluations within PRIs are used in relation to organisation-wide tensions will be discussed in the following sections.

2.2.2 Tensions within Research Institutions

Compared with the discussions at research systems levels, the governance tensions within PRIs are understudied by literature. Internal organisational governance of PRIs and their internal tensions are still considered a black box. To open such a black box, this study proposes **an analytical model with three organisational levels** involving various stakeholders. Basically, the multiple stakeholders of PRIs, presented in section 2.1.2, could be positioned in the three-level structure according to their interests and power. In this sense, inside PRIs, the key stakeholders are categorised as central decision-maker and supervisory bodies at the macro level (**L1**), that govern R&D activities and personnel as well as interact with external stakeholders; research institutes are found at the meso level (**L2**) under the organisational umbrella and comply with overarching policies and conduct complex R&D activities; and individual researchers are located at the micro level (**L3**), and participate in operational activities. These three levels of analysis (macro, meso, and micro) have been learnt from much governance-related literature, such as the methodology of political and social sciences (e.g. Jepperson & Meyer, 2011), the disciplinary governance forms of research organisations (Jansen, 2007), the aggregation machine of funding agencies (Rip, 2000, p. 467), and multi-level evaluation of scientific research (Hage et al., 2007).

The tensions between internal and external stakeholders of PRIs on organisational governance activities are attributed to their conflicting interests, the heterogeneity of assumptions and expectations regarding the contribution of PRIs to public research policy (Golden-Biddle & Rao, 1997). To justify public funding, the output expectations must be integrated by organisational governance to match all the external interests—scientific communities, national politics, the general public, the media, and the economy (Simon & Knie, 2013). Various internal stakeholders also have asymmetric information and heterogeneous perspectives about heterogeneous R&D activities. For example, there is no common ground for enabling key policy-makers, managers, and staff to come to an agreement about

which goals should have priority for improvement of a particular research programme (Palumbo & Hallett, 1993). Another example is proposed by Tampieri (2013) that the conflict of interests among project stakeholders leads to their adequate, effective, and efficient coordination in project implementation.

The complex characteristics of organisational governance at PRIs, on the three aspects—strategy, funding and operation discussed in section 2.1.1—are expected to perform differently at each organisational level. For example, organisational strategies concern the framing of policies to keep policy conflicts at manageable levels (Benz, 2007, p. 15), and mostly take place at L1. National or organisational policies on research funding and the relevant reforms are much more discussed at L1 than L2 and L3 (Grayson, 2015; Kupferschmidt & Vogel, 2014). The different legal forms of research institutes at L2 for non-university PRIs in Germany (Jansen, 2007, p. 234) have similar models concerning their internal organisational structures with the standard management model—the director or the executive board stand in the centre of the decision process with wide-ranging rights and are monitored by supervisory bodies at L1. To make governance effective, interaction and coordination of stakeholders across the three levels require interplay and transformation of tacit and codified forms of knowledge (Lundvall, 2002) on the three aspects, and collective actions of individual stakeholders (Benz, 2007, p. 15).

Furthermore, the three aspects of organisational governance at PRIs become the subjects and sources of internal stakeholders' tensions because of their diverging interests and power (Bryson et al., 2011). As Whitley and Gläser (2007, pp. 10–11) propose, the strategic capabilities of research organisations remain relatively restricted with respect to creating new or closing down old fields or units, which can conflict with advocates of aggressive strategies. The complex missions conflicting with scientific development of PRIs, as discussed in section 2.2.1, could be debated among internal stakeholders from different research fields and institutes. A financial conflict typically arises on *competitive grants* between various PRIs (Gläser, 2007, p. 249), and also between various fields and institutes within one PRI. Standardised operational targets within one PRI, such as active collaboration (Heinze & Kuhlmann, 2007) and good-quality education, are distributed to heterogeneous research fields and institutes, which may result in tensions. For example, Leisyte et al. (2009) propose that teaching and research are increasingly falling apart as two distinct activities in the context of university governance reforms; the balance between teaching and research can also be

targeted by PRIs. These hypothetical conflicts are still to be tested. We will come back to them, especially in Chapter 5.

Therefore, given the different roles of PRIs in their national and institutional contexts, **governance tensions of PRIs are fundamentally attributed to the stakeholders and their diverging interests in the governance aspects (strategy, funding, operation) and their powers at the particular organisational levels (L1, L2, L3).** The hypothetical patterns of governance tensions, both aspect-wise and level-wise, will be explored empirically in the case studies and responded to by the enriched conceptual framework in section 4.5.

To mitigate or overcome the stakeholders' tensions, both external and internal tensions, is proposed by this thesis to be an overarching target of organisational governance of PRIs. As Simon and Knie (2013, p. 408) propose, the objective of organisational governance is a perfect balance between external regulation (science policy) and internal self-governance of heterogeneous R&D activities. Literature on stakeholder analysis offers some solutions to facilitate satisfactory conflict resolution. For instance, committed group members attempt to resolve conflicts by actively communicating with each other about their controversial motives or ideologies and then by engaging certain stakeholders in collective negotiation (Golden-Biddle & Rao, 1997). Ultimately, a wide range of methods and procedures work for addressing conflicts, including but not limited to negotiation, mediation, diplomacy, and creative peace building (*ibid*). As will be explained further in this thesis, evaluation could be one of such methods. The next question is—can evaluation mechanisms work as useful instruments for PRIs to reflect, balance, or mitigate governance tensions, and how?

2.3 Evaluation Mechanisms as Governance Instruments

This thesis agrees with Jacob et al. (2015, p. 7) on that 'there are very few normative claims regarding how evaluation should be embedded in the architecture of governance'. Evaluation was originally conceived as an instrument to guide and improve projects, programmes and policies in the early 1960s but has increasingly grown to become an instrument integrated with the governance of organisations (Hasson, 2006, pp. 159–160). In addition, changes to the functional conditions for research have led to a growing interest in research evaluation since the 1990s and have provided an impetus for the application of relevant procedures (Kuhlmann, 2003, p. 134). Therefore, the question arises—does research

evaluation of, for, and at PRIs respond to governance requirements from various stakeholders as governance instruments of PRIs? As Peters and van Nispen (1998) claim, the understanding of *instrument* (policy instrument in their work) could be a part of a cognitive paradigm in an appropriate framework of ideas, routines, and values shared by stakeholders.

What are evaluation mechanisms? Why are they needed and how are they used as governance instruments for PRIs? These questions are discussed in this section. As introduced, evaluation mechanisms are defined by this thesis as ***the ways the whole evaluation system is constructed by all individual evaluation activities and operate vis-à-vis the complex institutional environment and stakeholders***. This definition could be applicable for both national and institutional levels. Although the existing literature focuses on national or EU-like systems, this thesis aims to find more evidence at the institutional level for the core idea that evaluation mechanisms help PRIs to respond to requirements of organisational governance. Section 2.3.1 presents the systemic nature of evaluation mechanisms of both research systems and organisations. Section 2.2.2 discusses the relevant literature on how evaluation mechanisms can be institutionalised for governance purposes. Organisational use of evaluation mechanisms as a balancing medium for reflecting and mitigating governance tensions is discussed in section 2.2.3. All the evidence in these three sections is collected from rich but disconnected evaluation studies.

2.3.1 The Systemic Nature of Evaluation Mechanisms

According to Arnold (2004, p. 3), ‘a systems world needs systems evaluations’ for the scope and practice of research evaluation where reality, theory and the needs of users have moved beyond the lower levels of individual projects and programmes towards a systems perspective. As a result of the systems perspective, the evaluation scope increases with more general monitoring and strategic impact and the evaluation approach becomes aggregated (Arnold, 2004, pp. 15–16). Mark et al. (1999, p. 177) propose an integrative evaluation framework that helps evaluators move beyond the segmentation of evaluation theories and approaches. Such an overarching framework that includes diverse evaluation approaches has a similar target with our defined *evaluation mechanisms*; because both of them present explicitly the strengths, weaknesses, and trade-offs of choosing one approach or the other, and the choices can be better communicated vis-à-vis the complex environment and all the involved stakeholders (*ibid*, p. 179). Therefore,

the **systemic nature of evaluation mechanisms** in this thesis is interpreted as the reason why they could be used as instruments for both governance of various PRIs within one system and internal governance within one PRI.

As Mark et al. (1999, p. 177) summarise, numerous evaluation theories—theory-driven, responsive, summative, formative, empowerment, goal-free, investigatory, outcome, four-generation, participatory, utilisation-focused, et al.,—provide flexible evaluators with a rich menu of approaches, most of which are claimed by their advocates to be distinctive. However, an integration of the theories and approaches is emerging. For example, there is a long history of calls for integration of quantitative and qualitative paradigms (Mark et al., 1999, p. 178). The classic method of quality-oriented peer review is slowly but steadily combining qualitative judgements with quantitative measurement of scientific productivity (science statistics, research mapping, bibliometric indicators and citation studies) (Hansson, 2006, p. 164). Arnold (2004, p. 16) proposes that the traditional *peer review*, mostly used at the project level, is increasingly modified for use at higher levels (portfolio and system) to become comprehensive *expert review*. The evaluation focus changing from lower levels (project, programme) to higher levels of aggregation (portfolio, system) needs a balance of appropriate evaluation methods (peer review, surveys, case studies, comparative studies, cluster analysis, etc.) (*ibid*, p. 15). Furthermore, the use of internal evaluations in research institutions has grown rapidly as a new instrument of including many of the above approaches (Hansson, 2006, p. 159).

The effect of evaluations advancing organisational governance has been overlooked (Simon, 2013) due to the lack of systemic consideration of institutional contexts in evaluation use (Edler, 2012; Højlund, 2014; Raina, 2003). To make the governance question visible, Hasson (2006, p. 169) emphasises the necessity to understand research evaluation in the context of research organisations. This argument is also reflected by the literature discussing the governance of public sciences in universities in different countries (Whitley and Gläser, 2007). For PRI studies in this thesis, both national and institutional contexts are considered carefully.

At national levels, the NIS and NPM, discussed in section 2.2.1, justify a systematic evaluation for public policy interventions about governance of state-funded science involving various PRIs (Arnold, 2004, p. 10). To allocate public resources in a less

wasteful way, an insistence on efficiency and relevance of public research has been an essential element of legitimization of research policy since the late 1970's (Schiene & Schimank, 2007). National metrics-based research assessment can be used to determine the allocation of funding among universities and research institutions (Geuna & Martin, 2003; Good et al., 2015; Whitley & Gläser, 2007). As Whitley (2007, pp. 9–10) claims, there are two ways of using research performance assessment to decide the allocation of block funding to research institutions—strong or weak research evaluation systems. The former uses a strong evaluation system to push immediate and schematic financial consequences—good performers are rewarded and weak ones are punished (Campbell, 2005). The latter—a weak system, works instead through initiating organisational learning processes, and it will be decided case by case whether weak or average performers are given a chance to improve (Schiene & Schimank, 2007). The two research evaluation systems are suitable for different purposes: Regulatory science policy needs immediate financial decisions while curiosity-driven academic science prefers scientific improvement based on block funding. If a systemic evaluation mechanism includes the above two evaluation systems, the conflict between the two sciences, as discussed in section 2.2.1, could be ideally mediated by different governance arrangements for the different sciences within one research landscape.

Indeed, new tendencies and perspectives of public research governance drive systemic uses of different evaluations. For example, evaluations of public research organisations as a policy tool provide data for benchmarking national performance (OECD, 2006). For a national research system, to evaluate the cumulative effects requires an updated and contemporary national equilibrium that sews together the disparate evaluation evidence (Feller, 2007). The key is to coordinate multiple stakeholders' activities as recommended by *the UK Metrics Report 2015* (Wilsdon, 2015). A *Clearing House* was proposed by Kuhlmann & Edler (1998) as a communication hub within Germany's landscape allowing for open access to evaluation results. Ideally a well-founded and transparent evaluation hub would help improve the use of evaluation information individually and collectively. However, the *Clearing House* approach was not put into practice, because PRIs required confidentiality of their evaluation results, to be used for their own strategic purposes only, preventing excessive interference from outside.

For research institutions, evaluation mechanisms are supposed to have systemic consideration of multiple evaluation activities and respond to 'the forefront of

organisational thinking and behaviour' (Sanders, 2003, p. 318). PRIs' subsequent empirical reports on intensive and increasing evaluation activities prove that evaluation is becoming a taken-for-granted aspect of public administration and organisational development (Hansen 2003; Jacob et al., 2015). Individual evaluation subjects in PRIs (programmes, projects, fields, institutes, and scientists) serve a wide range of designated purposes and are implemented at different organisational levels (L1, L2, L3).

No individual evaluation activity can answer all potential governance questions, of course (Dahler–Larsen, 2011). According to Hansson (2006, p. 163), the increasing integration of evaluations is a forceful demonstration of how evaluation mechanisms become an integrated part of organisational governance. For example, the combination of qualitative peer review and quantitative research measurement (Hansson, 2006, p. 164) in PRIs can provide evidence to benchmark heterogeneous R&D activities across various disciplines, fields, and institutes. The goals of both summative evaluation, focusing on products' efficacy, and formative evaluation, aiming at process improvement, are included in overarching evaluation mechanisms (Fuchs et al., 1993; Scriven, 1996). The concept of *evaluation governance* is used to justify evaluation's impact on governance with a condition that evaluative information has governance initiatives and is anchored in systematic assessment rather than scattered subjects (Hansson, 2013, p. 47).

Furthermore, the systemic nature of evaluation mechanisms requires participation of multiple stakeholders involved in various evaluation activities. Any evaluation that fails to attend to key stakeholders would, as a consequence, be inaccurate, insensitive, and insufficient to make needed improvements and would lead affected leadership groups to avoid evaluation in the future (Bryson et al., 2011, p. 3). To prevent that, negotiation among key stakeholders (evaluators, policy-makers, managers, and researchers) is widely encouraged (Kuhlmann, 1999) to ensure that the multiple evaluation activities conform to the possible conflicting expectations of stakeholders. Based on that, how evaluation mechanisms involving multiple stakeholders are used as a moderation medium and then facilitate organisational learning will be discussed in sections 2.2.3 and 2.3.2, respectively.

In a nutshell, the systemic nature of evaluation mechanisms for both research systems and PRIs is assumed to help to respond to the complex governance requirements from multiple stakeholders. How this assumption works in the three

PRI cases will be discussed empirically in Chapter 6. As Arnold (2004, p. 6) claims, such a systems perspective of evaluation evolves along with changes of knowledge flows and the embedded innovation systems. It is agreed by this thesis and will be explained as institutionalisation of evaluation in the next section.

2.3.2 Institutionalisation of Evaluation Mechanisms

Jacob et al. (2015, p. 7) hold that evaluation can follow various designs and be embedded in different forms of institutionalisation in different sectors and on different levels. The institutionalisations of evaluation start at different points in time and proceed with varying speeds (Gläser, 2007, p. 245). In fact, there is **much less literature discussing institutionalisation of evaluation mechanisms at institutional levels than at national or system levels**. Usually *evaluation systems* are used in the literature at both levels. For instance, according to Whitley (2007, p. 5), ‘the institutionalisation of different kinds of research evaluation systems in many national academic systems is an important component of more general changes in the governance of the public sciences’.

Several efforts have been made by scholars and research evaluation experts to characterise the institutionalisation of evaluation systems at national or system levels. Gläser (2007, p. 245) demonstrates the institutionalisation processes of research evaluation systems, focusing on university systems—emerging from different initial situations and become institutionalised at the state level in Germany and Spain, at the national level in Spain, the Netherlands, Japan, Sweden, and Australia. Georghiou (2003, pp. 76–77) studies a diversity of evaluation frameworks of research and innovation policy and practices at the pan-Europe level, and finds increased cooperation and interaction between national research policies likely to create some degree of convergence. Georghiou and Kuhlmann (2002, p. 192) discuss conceptually the evaluation of the socio-economic impacts of public research policies in the European research area as a useful policy instrument. For doing so, a toolbox of evaluation methodologies has been provided by many experts, and include econometric models, cost-benefit analysis, expert panels, case studies, network analysis, technology foresight, benchmarking, and so on (Polt et al., 2002). Polt et al. (2002, p. 248) believe these evaluation tools have expanded from providing a means for quantification of policy impacts towards facilitating mutual learning from past experiences, supporting mediation, decision-making and policy strategy definition.

Some countries have national laws on research evaluation, a kind of institutionalisation of evaluation mechanisms. For instance, in South Korea, there is the 'law on the evaluation and management of performance of national R&D programmes' (OECD, 2014, p. 50). According to Byung et al. (2010, p. 120), the government-funded research institutes (GRIs) evaluation system under the National Research Councils (NRCS) in South Korea had a significantly positive influence on the acceptability of the evaluation results and the performance of these GRIs. In Germany, there is no similar law but the German Council of Science and Humanities (*Wissenschaftsrat*, an organisation maintained by all key stakeholders in the national research system) provides advice to the German federal government and the state (Länder) governments on the structure and development of public research, including many evaluations of institutes, programmes, disciplines, universities, etc. This particular organisation and its impact will be discussed in the German case studies in section 4.1.

Kuhlmann and Heinze (2004) analysed empirically the institutional setting of research evaluation in Germany, suggesting the establishment of an evaluation *Clearing House*, as discussed in the previous section. According to Cozzens (2003, pp. 62–63), a new framework for evaluating research in the US demands increased skills and knowledge beyond identifying and measuring the benefits of public budgets required by public accountability. In China, the concept of research evaluation has gained increasing attention since the 1990s as a tool of R&D 'administrations' rather than 'governance' (Fang & Tian, 2012, p. 20). When Chinese science became subject to massive transformation with huge investments, the focus and challenges of R&D spending and evaluations were debated by international scientific communities, for example in journals like *Nature* and *Science* (Cyranoski, 2014; Grayson, 2015; Michelle 2015; Sun & Cao, 2014). According to them, China's current research evaluation is developing more like the European model to strengthen S&T with stable institutional finance, rather than the American mode, which demands competitions for funding. However, the thesis finds contrary evidence to be discussed in Chapter 4.

At the institutional level, scholars discussing research evaluation systems so far focused on relatively passive responses of research organisations towards institutionalisation of national evaluation systems, mostly evaluation procedures. For example, as Whitley (2007, p. 9) presents, strong research evaluation systems institutionalise public assessments of the research quality of individual units,

including universities and PRIs, on a regular basis according to highly formalised rules and procedures. This thesis looks actively at the need to institutionalise evaluation mechanisms within research institutions focusing on their own organisational governance rather than national requirements. This thesis interprets the institutionalisation of evaluation at PRIs as ***the process of professionalising and systemising various individual evaluation activities with consideration of and uses for organisational governance***. In this sense, it is proposed that PRIs' individual evaluation activities institutionalised to become an integrated mechanism to increasingly respond to organisational governance involving various stakeholders. To test this, it is important to observe the dynamics of both governance requirements and evaluation subjects and scopes within PRIs.

As Polt et al. (2002, p. 15) propose, as both a policy- and practice-driven approach, evaluation has co-evolved with the experienced developments. Inspired by current literature (Borrás & Højlund, 2014; Dimaggio & Powell, 1983; Henry & Mark, 2003; Højlund, 2014; Meyer & Rowan, 1977), evaluation mechanisms are assumed to be institutionalised by the spread of **both procedures and cultures**. Required by NPM, as discussed in section 2.2.1, the main objective of highly formalised research evaluation is considered by Power (1997, p. 97) to challenge the power and discretion of relatively autonomous research organisations by making their performance more publicly accountable. For example, spreading procedures of research evaluation in universities is driven by political urges and budgetary pressures in many countries, such as the UK, the Netherlands, and Germany (Geuna & Martin, 2003). With the assistance of national interventions, a combination of regulations of evaluation activities has arisen to ensure the continued diffusion of evaluation procedures (Coglianese, 2012). As Arnold (2004, p. 5) suggests, procedures as well as results and impacts of individual evaluations (projects and programmes) are gradually systemised to support higher-level synthetic decision-making. Such a process is considered by this thesis as the ideal model for institutionalisation of evaluation mechanisms.

To institutionalise procedures of an *overarching evaluation mechanism* within one PRI, this thesis proposes to systemise individual evaluation procedures to make them complement each other. Rather than only formalising standardised procedures, such as guidelines, research organisations are encouraged to focus on networking and communities between various evaluation practices (Hansson, 2006, p. 171). Højlund (2014, p. 26) emphasises the importance of conditioning factors in

organisational use of evaluations because many evaluations aiming at policy improvement rarely succeed in practice. Moreover, there could be many challenges to practicing the ideal institutionalisation considering multiple stakeholders. The definition of *evaluation mechanisms* is unlikely to be neutral and unproblematic as it is a part of the policy process of organisational governance. There are endogenous conflicts within each evaluation activity as it is an interest-led effort by a particular stakeholder (Dahler-Larsen, 2011), such as a personnel evaluation for appointment or promotion (Schruff, 2013), selection of independent review panels for one institute, and fairness and transparency of evaluating competitive proposals across institutes (Van der Meulen & Rip, 2000). These specific considerations within individual evaluations are not addressed by this thesis. To measure the institutionalisation of evaluation mechanisms at PRIs, valid, meaningful, reliable, and summative dimensions are expected by the author but are not found in literature.

According to Jacob et al. (2015, p. 6), the meaning of institutionalisation of evaluation can be proved by institutional characteristics of trends in evaluation culture. The spreading of evaluation culture to trust in evaluation mechanisms as governance instruments calls for its acceptance by various stakeholders who are willing to negotiate and cooperate in evaluation activities. According to Jacob et al. (2015, p. 8), there is no single way of measuring evaluation culture within one institution; changes in critical dimensions in several institutions need to be tracked. A mature evaluation culture is led by several paths including the diffusion of trajectories of evaluation capacity-building and the development of evaluation culture following the existing organisational culture (*ibid*). Whether institutionalisation of evaluation mechanisms within PRIs by the spread of both procedures and cultures help to respond to complex governance requirements and stakeholder tensions will be explored empirically in Chapter 6.

2.3.3 Organisational Use of Evaluation Mechanisms as a Balancing Medium

How evaluation mechanisms could be used as governance instruments to reflect and balance stakeholders' tensions is discussed in this section. Such a balancing medium is inspired by the work of Kuhlmann (1999, p. 8)—S&T policy evaluation is used as a medium for the *moderation* of struggles, controversies, and negotiations of multiple stakeholders. The perspective of various stakeholders' challenges

conventional evaluation procedures but matches the systemic nature of evaluation mechanisms defined in this thesis. Does a similar ‘moderation’ medium, we call ‘balancing’, exist at PRIs?

In Kuhlmann’s paper, S&T policy-making faces by both conflicts of stakeholders’ interest and policy controversies across the fields of science, policy and industry (Hart & Kleiboer, 1995, p. 96). A successful moderation of generally cooperative strategies of stakeholders who are pursuing rival interests requires knowledge about the potential outcomes of different game strategies (Kuhlmann, 1999, p. 135). Moderation of S&T policy-shaping and decision-making processes becomes possible in the negotiating system of stakeholders if the rules and different stakeholders’ perspectives are known and influenced by moderators—policy administration actors (*ibid*). The key is whether evaluation procedures can be designed as a communication medium for moderation processes and put to practical use in the negotiating system (*ibid*). Therefore, for our study subject—PRIs that also participate in S&T policy-making at both national and institutional levels—**stakeholders’ negotiation** is interpreted by this thesis as the key to *mediate* and *mitigate* the tensions that were proposed in section 2.2. Therefore, the *balancing* role of evaluation mechanisms in this thesis is interpreted to include both the *mediation* process and mitigation effect of stakeholders’ tensions. Similar to S&T policy moderation, negotiating conditions, rules, and relevant knowledge must be considered in the balancing process of evaluation mechanisms at PRIs.

There are two widely accepted evaluation theories involving stakeholders’ communication and negotiation. The first is the theory of *Fourth Generation Evaluation* (Guba & Lincoln, 1989). This theory intends to define a mature evaluation approach that moves beyond previous orientations of measurement and judgement but develops to a new level with key dynamics around stakeholders’ negotiation including political, social, cultural, and contextual elements (*ibid*, pp. 8–13). Meaningful constructions of this theory are created for an interactive and learning process involving various stakeholders. The other theory is *Empowerment Evaluation* (Fetterman et al., 1996), which aims to foster improvement and self-determination by using self-evaluation and reflection in terms of evaluation concepts, techniques, and findings. It is widely used in programmes, organisations, and by individuals where communities of practice and a culture of learning are built among stakeholders (*ibid*, pp. 4–6). In this democratic process, any evaluated individuals or groups empower themselves, often with assistance and coaching, to

actively collaborate with others (*ibid*, pp. 8–10). According to Röbbcke and Simon (2002), self-evaluation is emphasised by organisations to strengthen their capacity to cope with stakeholders' interests and tensions. To acquire that capacity, PRIs need a self-critical approach and processes of self-reflection and self-assurance that are understood as a necessary part of research processes rather than as a state of emergency (*ibid*).

The above two evaluation theories, developed over two decades, share some features and elements with the *evaluation mechanisms* in this thesis—evaluation operating vis-à-vis the complex institutional environment and stakeholders. The shared key elements—stakeholders negotiation and learning, will be tested empirically in the PRIs' evaluation practices in Chapter 6. How stakeholders' interactions and negotiations contribute to organisational learning will be discussed in the next section. Here, the key question is how evaluation mechanisms can be designed as a communication medium for stakeholders and put to practical use in the negotiation of their tensions?

From a participatory perspective, evaluation should go beyond measurement to enable 'intelligent negotiations' among stakeholders who pursue different interests and strategies (Kuhlmann, 1999, p. 136). According to Röbbcke and Simon (2002), the selection of participatory procedures of stakeholders greatly affects whether PRIs succeed at designing evaluation as a process preventing excessive interference from outside the scientific system. For instance, in Germany, the more the evaluation of science has become externalised as a systematic and continuous process, political stakeholders are more increasingly involved in evaluations, by suggesting guidelines and intervening in financial decisions concerning individual PRIs (*ibid*, p. 168). However, in German universities and some research institutions, the traditional view of good research is related to individual professors and their research unit (Schimank, 1994, pp. 40–42). These individual units are starting to be evaluated as part of a larger organisation, which still appears new and strange to some German professors who think their constitutionally guaranteed scientific freedom to be violated (*ibid*).

How can *intelligent negotiation* be undertaken between stakeholders, individual professors, and politicians in the above example, to balance their tension? According to Van der Knaap (2006), stakeholders' dialogues in evaluations should use information and insights from diverse stakeholders to explicitly question the

validity of governance objectives without compromising their value and significance altogether. Involvement and interactions of diverse stakeholders (evaluators, policy-makers, managers, and researchers) are presumed to ensure that the effect of evaluation develops in accordance with all their expectations. The role of evaluators is no longer restricted to scientific peer groups but includes a growing number of professional evaluators, consultants, and representatives from politics, academia, businesses and so on (Hansson, 2006).

Given various stakeholders' expectations vis-à-vis organisational governance and related tensions, the balancing medium of evaluation mechanisms of PRIs faces managing many sources of influence from various evaluation activities for balanced governance arrangements (Cruz-Castro & Sanz-Menendez, 2007, p. 207). For example, peer-review-based research evaluation focuses on experts' feedback and is used as an authoritative source for research institutes to justify and optimise their governance arrangements (Simon & Knie, 2013). Based on that, integrated recommendations from various evaluation activities are negotiated with more diverse stakeholders—funders, decision makers, supervisory bodies, managers, etc. about how organisational governance of the whole institution could be facilitated and how institutions' development can be supported in the research systems. The tensions proposed in section 2.2 can be mediated in the stakeholders' negotiations.

Stakeholder participation procedures in evaluation mechanisms include round tables and nationwide or organisation-wide debates (e.g. Felt & Fochler, 2008; Fischer, 2009; Jasanoff, 2003; Rayner, 2003). Before involving stakeholders in participatory procedures, a broader and inclusive approach to defining key stakeholders is needed to narrow the focus on the major stakeholders who can affect and be affected by evaluation processes and findings (Bryson & Patton, 2011). Unfortunately, the challenges of identifying and including stakeholders, capturing their perspectives, and embracing their concerns are always underappreciated in practice due to budget and time constraints (Bamberger, 1991). Return on evaluation investment, such as costs and benefits of involving stakeholders, is highly context-specific and path-dependent (Arnold, 2004, p. 14).

To sum up section 2.2, evaluation mechanisms of PRIs are used as governance instruments because they can be designed with a systemic nature and institutionalised to respond to governance requirements. To reflect and mitigate governance tensions, involvement and negotiation of key stakeholders can be

designed and practised (to be tested) in procedures of evaluation mechanisms as a balancing medium. In the processes of evaluation mechanisms functioning for organisational governance, organisational learning is also facilitated, which will be discussed in the next section.

2.4 Evaluation Mechanisms Facilitating Organisational Learning

Learning is widely accepted as the most important process in innovation activities (Johnson, 2002) and is increasingly inserted into various forms of innovation networks involving PRIs. Policy learning refers to the integration and processing of knowledge leading towards action (Hall, 1993). Organisational learning is identified as one of the most important perspectives with which to approach organisational change (March, 1981). As Levitt and March (1988) propose, organisations integrate events and knowledge into their organisational routines, and adapt their behaviour based on their targets and the overall contexts where they evolve and react. As Lundvall (1992) proposes, learning, searching, and exploring in innovation systems are the sources of technical innovation that create the basis for new paradigms of research itself and also governance of research.

Why and how organisations learn is emphasised by literature across several research fields. Science itself is changing and learning all the time, therefore, the relevant governance and evaluation of scientific work are required to keep learning correspondingly (Borrás & Højlund, 2014). What does *organisational learning* mean? In research on public organisations, learning has been studied as the source of a number of changes in values, attitudes, structural and functional adjustments, and behaviours that explain the nature and role of the learning process (Barrados & Mayne, 2003). Based on that, organisational learning for PRIs is interpreted by this thesis as integrating knowledge into organisational routines and adapting organisational behaviour and changes in the function of the organisational missions.

What systemic nature evaluation mechanisms have and how it can be used as a balancing medium to facilitate organisational learning of PRIs will be discussed in this section. Two functional poles of expectations of research evaluation can be identified (Kuhlmann, 2003, p. 141): a) the summative function measures performance and provides legitimisation; b) the formative function is utilised as a *learning medium* where findings of the measures are considered intelligent information for current or future initiatives (*ibid*). Based on both poles, this thesis proposes **two hypothetical ways of the *learning medium* of evaluation**

mechanisms: a) evaluation findings (measurement and legitimation) produce knowledge bases for organisational governance; b) evaluation processes involving stakeholders' interactions and negotiations facilitate organisational learning. The two ways will be tested empirically in section 6.4.

2.4.1 Knowledge Bases Used by Organisational Governance

What kinds of knowledge bases produced from evaluation findings and used by organisational governance are discussed in this section. Expert knowledge is considered by literature, such as Hoppe (2010), as a key element in policy-making and policy learning. Knowledge utilisation studies emphasise the importance of communities' cultural and behavioural differences as well as participation of policy-makers (Padilla, 2016). Evaluation research is identified as one of the mainstream research areas that contributes knowledge to be utilised by policy-making (Hutchinson, 1995). Likewise, various channels and different types of knowledge uses are covered by the evaluation utilisation literature (Rich, 1997). The utilisation and usefulness of evaluation processes and outcomes are addressed by many evaluation studies, such as Botcheva et al. (2002).

The different aspects of governance (strategy, funding, operation) at PRIs require different kinds of knowledge. As Hasson (2006, p. 157) suggests, various evaluation activities produce a set of data and the long-term impact of the use of these data on organisational activities. To be specific, various evaluation activities at PRIs explicitly contribute their individual objectives and uses, for example, public accountability (Cozzens, 2003), programme improvement (Fetterman, 1996), decisions on allocation of research funding (Whitley, 2007; Good et al., 2015) and of other resources in internal governance processes (Röbbecke & Simon, 2002), and research policy legitimacy and interventions (Amanatidou et al., 2014). Distributed and heterogeneous 'evaluative information and services' (Heinze & Kuhlmann, 2004, p. 60) can be produced, transformed, negotiated, interpreted, and reported (Kuhlmann & Edler, 1998) across various organisational units and levels. Accelerated by the systemic use of individual evaluations, evaluation mechanisms are increasingly considered as a part of the strategy and culture of the organisations by both the organisation at L1 and its members at L2 and L3 (Hasson, 2006, p. 159). However, the literature very rarely discusses the impact of the uses on the behaviour and activity of organisational members, groups, and individuals (*ibid*), which will be explored empirically in Chapter 6.

Based on that, this thesis proposes that evaluation mechanisms of PRIs can accumulate, analyse, and aggregate compatible *evaluation information and services* as regular evidence for governance decision-making. The information flows into a dense network enabling PRIs to strategically manage the continuing development of the individual units and of the whole institutions. The consideration of who is using what knowledge for governance to better understand evaluation uses is highlighted in this section. Asheim et al. (2011, p. 168) propose three types of differentiated knowledge bases types (*analytical, synthetic, symbolic*) that characterise different types of knowledge as both outputs and inputs of innovation activities. These three knowledge bases were originally used by the Organisation for Economic Co-operation and Development (OECD) to classify industries and have been used by this study to categorise the knowledge outputs of evaluation mechanisms as inputs for PRIs' governance. The way these three knowledge bases are categorised can be seen to reflect the different essences of various evaluation findings at PRIs.

Both *analytical* and *synthetic* knowledge bases are mainly produced from the innovation activities themselves, while the *symbolic* one refers to a verification of professional or cultural sources. The *analytical* knowledge performs as new scientific knowledge depending on formal models and codification within professional R&D activities (Asheim et al., 2011, p. 169). The *synthetic* knowledge refers to the application, 'novel combinations', or aggregation of existing knowledge for their integrated uses in innovation activities (*ibid*). The *symbolic* knowledge is more dedicated to the creation of new ideas and designs than to an actual physical process, which requires abilities and creativity of professional originators and leaders (Ross et al., 2007). The innovation activities—often economic activities in Asheim's work—are interpreted by this thesis as scientific activities and their evaluation and governance at PRIs.

When applied in this study, the *analytical* knowledge generated from individual research evaluations explicitly brings about bases for governance decision-making on the evaluated target. This way is assumed to be suitable for most types of research evaluation at PRIs. Furthermore, the analytical knowledge inputs for governance need codified knowledge outputs from professional research assessment using scientific principles and methods. These assessment activities require a high level of expertise and analytical capabilities of evaluators. This type of knowledge that is created in forms of scientific discoveries and technological inventions is much more obvious than the other two types.

In contrast, the *synthetic* knowledge is assumed in this study to be generated from comparative evaluations of several subjects, like benchmarking. This type of knowledge is more frequently produced in the evaluation of strategic and applied research that aim to solve specific problems across units or fields, than in evaluation of unique topics in fundamental research. Additionally, the *synthetic* knowledge is often gained through practical work and experience of learning-by-doing skills (Asheim et al., 2011, p. 169). In this sense, the *synthetic* knowledge generated from various evaluation practices within evaluation mechanisms leads to an incremental organisational learning through modification and improvement of the existing evaluation processes and uses.

Symbolic knowledge bases in this study are interpreted as the creation of significance beyond evaluation activities themselves but focusing on the involved people like evaluators whose creative and interpretive skills are required. This type of knowledge is more important when justifying *who* rather than *how* to conduct research evaluations, because the knowledge of professional people (evaluators, originators, policy-makers and collaborators) with specialisation and networking in communities are much more important than evaluation procedures (Hansen, 2013). In this sense, evaluators and other professional people rather than evaluation activities or findings contribute to, as well as take advantage of, such a knowledge base. Moreover, as Højlund (2014) supplements, evaluation results, typically evaluation reports, could be used symbolically as justification for the evaluated subjects. For each PRI, mixed knowledge bases can probably be generated and used for complex governance decision-making. All three knowledge types will be tested empirically in the case studies.

2.4.2 Organisational Learning Facilitated by Learning Evaluation

In principle, evaluation itself can be learning all the time. The concept of *learning evaluation* has been discussed by the literature focusing on continuous improvement by learning from evaluative recommendations and lessons (Edelenbos & Van Buuren, 2005), from stakeholders' dialogues and negotiation (the Fourth Generation Evaluation by Guba & Lincoln, 1989), and from the learning capability of stakeholders (Van der Knaap, 2006). Schön (1983) defines *learning evaluation* as a process of trying and adapting, acting and interpreting, learning-by-doing (Johnson, 2002) in a continuously changing environment in which both the evaluated and evaluators as reflective practitioners shape the evaluations in close

interaction and consultation. As Dahler–Larsen (2011, p. 226) proposes, the culture of learning evaluation is influenced by institutional models where ‘most evaluators subscribe to the positive vision of the learning organisation’. Based on the above discussion, *learning evaluation* depends on its working environment—in this thesis, the PRIs and the stakeholders.

The followed question arises: Can PRIs’ *organisational learning* be facilitated by *learning evaluation*, and how? According to Kuhlmann (2015, p. 8), evaluation as a tool of strategic intelligence facilitates single- and double-loop policy learning at organisational levels. Single-loop learning helps to keep organisational performance within the range set by organisational norms (Argyris, 1977). Double-loop learning concerns incompatible organisational norms by setting new priorities and weightings of norms, or by restricting the norms together with organisational strategies (Kuhlmann, 2015, p. 9). Consequently, learning organisations have a rapid rate of changes in relation to both existing performance and competence, and creating new necessary ones (Dahler–Larsen, 2011). To support such changes, evaluation findings, such as knowledge bases discussed in the previous section, and learning stakeholders’ negotiations can be used for PRIs to analyse strengths and weaknesses of organisational changes with perspectives for future development, such as the ‘Detector–Director–Effector learning loop’ from knowledge to action adaptable to changing environments in organisations (Dahler–Larsen, 2011, p. 45).

How can stakeholders’ interactions and negotiations be learnt by organisations? This thesis proposes that organisational governance of PRIs learn to become increasingly responsive to diverse stakeholders, including their governance requirements as well as tensions, from stakeholders’ interactions and negotiations in the evaluation processes. According to Van der Knaap (2006), the quality of stakeholders’ dialogues should be raised to help organisations remain sufficiently responsive to external and internal changes and developments. Preskill and Torres (1999) believe organisational learning from stakeholders is characterised by a collective engagement with a focus on identifying and addressing the commonly held issues and initiatives of stakeholders. Therefore, organisational learning requires formal procedures and patterns of stakeholders’ interactions that consider participants’ previous knowledge and experiences and their group behaviours (*ibid*). The practical procedures and patterns in evaluation activities can be observed through required dialogues of stakeholders, such as meetings for discussing evaluation guidelines, programmes, or results.

Lundvall (1992) proposes four types of learning in innovation activities in terms of increasing interaction between people: imprinting, role learning, learning by feedback, and searching. This thesis focuses on *learning by feedback* since there is much stakeholders' feedback in evaluation activities (Dahler–Larsen, 2011). Beyond gathering and analysing feedback on evaluation results, organisational learning is assumed by this thesis to come from stakeholders' feedback at each step throughout the evaluation processes (Rebien, 1996). Simon and Knie (2013) emphasise the form of communication before, after, and during evaluation processes: which status groups are included, and in what way, and in what manner feedback occurs. For example, on-site evaluations at PRIs can involve dialogues of diverse stakeholders such as scientists, international experts, local politicians, and businessmen. Their feedback on guidelines, implementation, and uses of evaluation can enlarge the impact of evaluation, which influences organisational learning over time and space.

Furthermore, Guerrieri and Tylecote (1997) find three kinds of stakeholder interaction in innovation systems—functional, vertical, and external—that can be applied in the study of PRIs. A functional interaction could refer to the horizontal interaction between units like institutes at the same levels within PRIs. A vertical interaction means interaction up and down, across the different levels. External interaction takes place between one PRI with other institutions. Therefore, the regularities, practices, and people's acceptance of stakeholders' interactions and negotiations within both the individual institutions and the communities between institutions are considered by this study. Fisher and Schuurbijs (2013) emphasise that interactions can be effective only when they are structured, sustained, embedded, and open-minded. These elements are considered in the empirical chapters when discussing stakeholders' interaction at the PRI cases.

The discussions, about ways of organisational learning, can turn normative in cases where successful PRIs are capable of responding to various stakeholders. When prestigious organisations do so successfully, numerous others may copy the mechanism (Dahler–Larsen, 2011). It has to be noted that organisational learning of PRIs includes, but is not limited to, the instrument of evaluation mechanisms, not at least since they have their own tensions. According to Van der Knaap (2006), evaluators face dilemmas between justice to societal complexity and maintaining scientific focus, between providing practical and useful tips and maintaining sufficient independence and authority, and between improvement and innovative

learning. Edelenbos and Van Buuren (2005) compare the role conflicts of evaluators serving several target groups simultaneously (adviser, assessor, facilitator, coach, negotiator bridging opposition, etc.) with multi-voice situations. Laudel and Gläser (2006) point to the tensions between the simplifying, standardising approaches of evaluations and the complex nature of communication practices in science in various performance channels (journal, book, conference paper, report, and non-publication). The above internal tensions of evaluations and the possible consequences or impact will not be further addressed by this thesis.

2.5 Research Question, Sub-questions and Propositions

The overarching research question remains: ***How do evaluation mechanisms of PRIs play a balancing role and help to mitigate tensions in organisational governance?*** Four sub-questions can be derived from the above conceptual discussions to be investigated in the subsequent case studies:

- ① Which specific factors in national and institutional contexts cause the tensions of organisational governance for PRIs?
- ② Which stakeholders are involved in tensions of organisational governance and why?
- ③ What are the balancing processes of evaluation mechanisms reflecting and mitigating governance tensions?
- ④ How do evaluation mechanisms facilitate organisational learning for the mitigation of governance conflicts?

These are four propositions reflecting the sub-questions:

P1: Tensions of organisational governance at PRIs relate to the three governance aspects (strategy, funding, operation) and to various perspectives of the involved stakeholders in national and institutional contexts.

P2: Governance tensions involve central supervisory bodies at the macro-level (L1), research institutes at the meso-level (L2), and individual researchers at the micro-level (L3) because of their diverging interests in the governance aspects and power at particular levels.

P3: Governance tensions can be reflected in the institutionalisation of evaluation mechanisms and mitigated by interaction and negotiation among various stakeholders.

P4: Evaluation mechanisms facilitate organisational learning by providing knowledge bases of evaluation findings and stakeholders' interactions and negotiations in evaluation processes.

2.6 A Basic Framework on the Role of Evaluation Mechanisms in Organisational Governance

Based on the literature reviewed, a basic conceptual framework can be sketched with two core concepts, *organisational governance* and *evaluation mechanisms*, as well as their co-evolution. Figure 2.1 presents this heuristic. To answer the overarching research question, the **balancing** role of evaluation mechanisms is reflected by the two core steps of the whole circle in Figure 2.1: a) governance tensions involving various stakeholders justify the purposes of evaluation mechanisms; b) evaluation mechanisms respond to and mediate tensions and facilitate organisational learning for the mitigated organisational governance. Therefore, the *balancing role* of evaluation mechanisms is interpreted in this thesis to include both the mediation process and mitigation effect of stakeholders' tensions, to keep organisational governance of PRIs in equilibrium.

In Figure 2.1, *organisational governance* is the independent variable in this study. Its essence and tensions were discussed in section 2.1, considering the embedded national and institutional contexts. Overall, the national public R&D landscape requires justification of public resources invested in PRIs; and institutional contexts shape organisational governance concerning the three key aspects (strategy, funding, operation) involving multiple stakeholders across three organisational levels (L1, L2, L3). *Tensions*, as a dependent variable of stakeholders' interests in the governance aspects and stakeholders' power at organisational levels, are assumed by this thesis to justify evaluation mechanisms as governance instruments. Next, *evaluation mechanisms* are influenced by, and also respond to, governance tensions in certain ways, such as institutionalisation of evaluation procedures and culture. Consequently, the balanced governance results, as the last dependent variable, are in turn facilitated by organisational *learning* from evaluation mechanisms. Here, the intermediate impact factors—governance tensions and

organisational learning—are not completely independent from organisational governance and evaluation mechanisms.

Figure 2.1 A basic framework on the role of evaluation mechanisms in organisational governance

Source: the author, 2015

Some gaps and opportunities of this study can be seen from the basic conceptual framework (Figure 2.1). There are still empty bubbles around the mutual impacts between organisational governance and evaluation mechanisms, which refer to the uncertainties of specific reasons, performances, and approaches of the *governance tensions* and *organisational learning*. The existing literature discussed in this chapter offered some clues. The sub-questions breaking down the overarching research questions and the relevant propositions will be explored empirically in the case studies. Consequently, those empty bubbles questioning about why and how the *governance tensions* and *organisational learning* take place will be presented in further refinement of the conceptual framework.

The following case studies will provide evidence to enrich this basic conceptual framework in an abductive way. The enriched conceptual framework (in section 4.5) is intended to position organisational governance in national and institutional contexts after confirming the exact influences of various stakeholders involved in these two contexts. After that, discussions on which stakeholders cause which kind of governance tensions will become possible. Then responses of evaluation mechanisms towards the governance tensions will be analysed. The final conceptualisation steps (in section 7.1.5) should help to outline clearly the contexts, essences, performances, and influences underlying the *governance tensions* and *organisational learning* in Figure 2.1 so as to make the whole *balancing circle* complete.

To sum up this chapter, the conceptual discussions based on the existing literature focus on the concepts of *organisational governance* and *evaluation mechanisms* as well as their relationships. The conceptual discussions help to frame the research sub-questions and propositions that are derived from the overarching research question. A basic framework has been established to explore evidence to validate the relationships between the core concepts by testing the propositions in case studies. This basic conceptual framework (Figure 2.1) will be developed in an abductive way into the enriched and final competed frameworks, with more and more elements added to respond to the research sub-questions. The next chapter will present the research approach to operationalise the key concepts and their relationship discussed in this chapter.

3 Research Approach

How can the concepts in the basic conceptual framework proposed (Figure 2.1) be operationalised in the cases of Max Planck Society (MPG), Helmholtz Association (HGF) and the Chinese Academy of Sciences (CAS)? How can the propositions be tested? This chapter shows the specific steps to do so. Section 3.1 introduces the approaches to identify the stakeholders in relation to the three PRIs at each organisational level. Section 3.2 explains how to operationalise the key concepts for planning the fieldwork to collect interview data. Overall, the empirical data comes from academic literature, practical documents (official websites, annual reports, policy documents), and 57 interactive interviews with the stakeholders. Interview guidelines for collecting data are presented in section 3.3. An overview of how the case evidence was analysed is shown in section 3.4. The validity of this research approach is illustrated at the end of this chapter.

3.1 Applying the Three-level Structure into the Case Studies

Table 3.1 presents the stakeholders of the three PRIs actually visited and interviewed in line with the three organisational levels (L1, L2, L3) discussed earlier. This structure guided the selection of subjects in the research design phase, which was in turn confirmed in practical interviews. Centred on the key concepts within the basic conceptual framework, the topics discussed at each organisational level concerned different interests and the power of the selected stakeholders.

Subjects	MPG	CAS	HGF
Macro L1: central decision-making and supervisory bodies	Administration Office of headquarters; Department of Research Policy and External Relations	Administrative headquarters; Evaluation centre;	Research Section of Head Offices
Meso L2: Research institutes or centres, (directors, secretary, administrators)	MPI A	CAS a	
	MPI B	CAS b	
		CAS c	HGF X
		CAS d	HGF Y
	CAS–MPG Partner Institute for Computational Biology (PICB)		
			the Helmholtz Beijing Office (HBO)
Micro L3: individuals	professors, associate professors, post-docs, and PhDs		

Table 3.1 Actual subjects visited and interviewed

Selection of the stakeholders related to the three PRIs at L1 and L2 was carefully considered. Yet, there was no special rule for selecting interviewees at L3 as long as they participated in evaluation activities and agreed to share individual experiences.

Selection of the Institutions at L1

Why were the three PRIs (MPG, HGF, CAS) selected as case studies? Their commonalities and differences of organisational governance were taken into account in terms of both quantitative performance, as shown in Table 3.2, and qualitative characteristics of strategy, funding and operation. First, both Germany and China have heavily invested in their national research systems that influence the PRIs' governance, which will be explained in section 4.1. Second, as introduced at the beginning of the thesis, the three PRIs have excellent international reputations according to the *Nature Publishing Index* (2013). Third, they all have developed for decades and been funded by both central and regional governments. Generally, the numbers of employees and annual funding levels of CAS represent those of MPG and HGF combined. Their different proportions of public block funding and project funding are closely related to their governance arrangements, which will be explained in Chapter 4. Fourth, the three-level organisational structure exists at the three PRIs and their respective tensions are related to such similar structures, which will be presented in Chapter 5. According to Table 3.2, all the three PRIs have large numbers of research institutes under their umbrellas at L2 and many individual researchers at L3. In fact, the three PRIs have active research collaboration at L2 and L3 and share joint research centres or institutes at L2, such as the CAS–MPG Partner Institute (PICB in Table 3.1). All the above similarities and connections provided rich evidence for the comparative analyses.

Institutions	Research institutes	Employees	Annual funding (billion euro)	Proportion of public block funding (%)	
				Central government	Regional government
MPG	82	21,640	1.93	85	
				50	50
HFG	90 (in 18 centres)	33,619	3.76	70	
				90	10
CAS	104	67,900	5.63	53	
				94	6

Table 3.2 General statistics of the three research institutions

Source: the 2013 annual reports of MPG, HGF, and CAS

Furthermore, the different characteristics of the organisational governance and evaluation mechanisms at the three PRIs provide a diversity of evidence to answer our overarching research question. For example, how their different research orientations (fundamental research at MPG, strategic research at HGF, and mixed research at CAS) influence their governance challenges, arrangements and evaluation mechanisms will be analysed in Chapter 4. Some similar and generalisable findings based on those differences and presented in Chapters 6 and 7, validated the selection of the institutions.

Additionally, both MPG and HGF have developed research evaluation involving all their institutes or centres at L2 over 15 years. Yet, CAS started to standardise research evaluation for all the institutes in recent years. CAS has been learning from experiences in other parts of the world, especially in Germany. The exchange of experiences and learning among the three institutions concerned both research itself and research evaluation, which provided comparative clues in our interviews. The interviewees chosen at L1 are considered as the most appropriate stakeholders, who could be approached by the author, to discuss the research questions and related topics from the perspective of the whole institutions.

Selection of Member Institutes at L2

As Table 3.1 shows, particular institutes or centres were selected to discuss our research questions and related topics, such as coordination of L1's governance and evaluation policies. To ensure the comparability, it was important to select the institutes from the three PRIs working within similar fields or on similar themes. For example, CAS institute c and HGF centre X both conduct health research. In particular, MPI A and CAS institute a both work on societal impacts of science and technology (S&T) and share a deeper understanding of evaluating S&T than other institutes working with natural sciences. The inclusion of the two joint units was very meaningful—PICB and HBO in Table 3.1—and provided much information on their experiences of two different systems.

The interviewees chosen at L2 were mostly institute (or centre) leaders (directors, managers, or deans) and evaluation coordinators from various departments within the institutes (or centres): departments of S&T Administration, of Communication, of Technology Licensing or Transfer, and of Human Resources and Education. The specific topics discussed at each level will be presented in section 3.3.

3.2 Operationalisation of the Key Concepts within the Case Studies

The key concepts of this thesis—organisational governance, tensions, evaluation mechanisms, learning and balancing—were operationalised in the case studies to explore evidence to answer the research questions and to test the propositions. In fact, the three cases (MPG, HGF, CAS) were first studied individually and then compared, which resulted in the synthetic analyses in Chapters 4–6.

Characteristics of Organisational Governance

The key governance aspects (strategy, funding, operation) were analysed before the interviews and based on the self-published resources of the three PRIs, such as official websites and annual reports, and the related academic literature. Then the analyses and preliminary findings were discussed in the interviews. Efforts were made to prevent the loss of objectivity when collecting data from the positive arguments in the PRIs' reports. The comparative Table 3.3 summarising the characteristics of the three PRIs was shown to the interviewees in advance as an entry point guiding the discussions on the strengths and weaknesses of their current governance, and what efforts were or could be made to develop the strengths and overcome the weaknesses.

	Strategy		Funding		Operation		
Institutions	①	②	③	④	⑤	⑥	⑦
MPG	Fundamental research						
HGF	Strategic research						
CAS	Mixed research						

Table 3.3 Governance characteristics of the three institutions

Specific governance characteristics were explored based on the following points:

- ① Organisational mission and research orientation
- ② Strategic plans and activities to achieve the missions
- ③ Proportion of public funding from central and regional governments, and proportion of institutional block funding and competitive project funding
- ④ Fund allocation and competition among member institutes or centres

- ⑤ Legal status and autonomy of institutes or centres
- ⑥ Management style: top-down or bottom-up
- ⑦ Other operation activities such as human resources, collaboration, and training.

All these points are analysed in Chapter 4 and the national and institutional contexts of the three PRIs are considered. In particular, the analyses provided a basis to understand how the characteristics of organisational governance were related to the stakeholders within those contexts and to their tensions.

Stakeholders and their Tensions

Who are the *stakeholders* of PRIs? Bryson et al. (2011) and Eden and Ackermann (1998) provide a useful method to analyse stakeholders according to their powers and interests in research management and in research evaluation. The roles of stakeholders may change dynamically over time or because of other factors (Bryson et al., 2011, pp. 5–9). Table 3.4 summarising the stakeholder categories of the three PRIs was prepared before interviews based on the literature. Section 4.4 presents a filled-in figure, following confirmation by interviewees. The use of Table 3.4 in interviews helped the interviewees to better understand our research interests and discuss their roles and also those of other stakeholders. The stakeholder literature helped with the selection of stakeholder interviewees and the preparation of distinct interview questions (in section 3.3). Since it was impossible to get access to every key stakeholder, it was important to infer the involvement and interaction among various stakeholders, based on the views of the interviewees at the PRIs.

Empirical evidence of *stakeholder tensions* was traced from the interviewees' divergent opinions, disagreements, and arguments concerning both governance and evaluation. The word *tension* sounds so negative that all the interviewees did not like to talk about it explicitly. Understandably, very few people interviewed actually complained about the PRIs where they worked. Rather than the existing tensions, some interviewees, mostly at L1 and partly at L2, preferred to talk about how they successfully prevented certain predicted problems that had already happened or were expected in the near future. In this sense, the concept of *tension* was interpreted by the author from the stakeholders' actions and efforts that aimed to prevent some negative consequences they anticipated. It was a process of probing, interpreting, and refining rather than finding out particular *tensions*. The

literature on governance tensions, as discussed in section 2.2, offered experiences and tips. Furthermore, some stakeholders were asked whether they agreed on the conflicting issues proposed by other interviewees or by academic literature, such as policy-driven science versus curiosity-driven science. Some stakeholders gave answers not only from their own perspectives at a particular level, but also reflecting other levels.

How are they constituted and what are their concerns?		Institutions	HFG	MPG	CAS
Stakeholders					
Internal stakeholders A	① Senate Committee				
	② Scientific Advisory Board				
	③ Evaluation Centre				
	④ Headquarters				
	⑤ Internal researchers				
	⑥ Member institutes or centres				
External stakeholders B	① Governments				
	② Ministries				
	③ Scientific community				
	④ Research Council				
	⑤ Industrial enterprises				
	⑥ Universities				
	⑦ Public				

Table 3.4 Stakeholders of the three institutions and their concerns

Evaluation Mechanisms

Evaluation mechanisms are defined by this thesis as to include all individual evaluation activities and operate vis-à-vis the complex institutional environment and stakeholders. This definition received more feedback from interviewees at top organisational levels than at lower ones. To make this definition clearer, Table 3.5

was prepared including preliminary findings on the subjects and features of the individual evaluation activities at the three PRIs. Then this table was showed to the interviewees to discuss specific evaluation activities they participated in. If the features of individual evaluation activities of one PRI were systemically organised to supplement each other, the systemic nature of an evaluation mechanism could be confirmed. For example, distinct evaluation activities focusing on different levels (system at L1, institutes at L2 and individual at L3) could constitute a comprehensive evaluation mechanism involving all the three levels.

<i>Subjects</i> <i>Features</i>	Project	Programme	Institute	Centre	Cluster	Researchers
Level						
Scale						
Timing						
Participation						
Use						
Impact						
Confidentiality						

Table 3.5 Evaluation activities composing evaluation mechanisms at the three PRIs

Organisational Learning

The essence of *organisational learning* as defined in section 2.4—integrating knowledge into organisational routines and adapting organisational behaviour and changes in function of the organisational missions—was inquired in the three cases. By reviewing the academic literature and policy documents concerning governance dynamics of MPG, HGF, and CAS in the past 15 to 20 years, the particular organisational behaviour and changes with respect to their evaluation activities were ascertained and discussed in the interviews. What kind of knowledge is produced from evaluation and how it is integrated into organisational routines were explored in the evaluation related documents and discussed in the interviews.

3.3 Interview Guidelines and Data Collection

Interview guidelines were prepared to efficiently collect interview data concerning the key concepts and their relationships proposed by this study. The interview guidelines made for the distinct organisational levels were shown to the interviewees in advance and included definitions of and questions about the key concepts. To ensure interviewees understood and were able to contribute

substantively to the overarching research question, related questions for other levels were also included. Actual conversation points were added and changed according to the responses of interviewees as the interviews progressed in this study (Yin, 2009). Although the precise wording varied from time to time and from interviewee to interviewee, an overview of the key questions is set out in Tables 3.6 to 3.8. Pro forma questions were prepared to encourage interactive conversations to dig deeper and reflect the interviewees' opinions regarding the role of evaluation in organisational governance.

About the characteristics of individual evaluation activities (see Table 3.5), how are they designed to account for governance requirements? For example, which evaluation is requested to produce evidence for strategic or financial decision-making?
Do the individual evaluation activities supplement each other to constitute an overarching evaluation mechanism with a systemic nature, in terms of subjects, levels, uses, etc. (see Table 3.5)?
How does the headquarters coordinate evaluation design, practices and uses with other stakeholders (see Table 3.4)?
How do evaluation results impact the stakeholders at each organisational level on the governance aspects of strategy, funding and operation?
Concerning governance dynamics and evaluation institutionalisation, do you think that evaluation can help to facilitate organisational governance, and how?

Table 3.6 Interview questions for L1

Some preliminary studies on the organisational governance and individual evaluation activities based on literature were discussed with the interviewees. Interview questions for stakeholders associated with central decision-making and supervisory bodies at L1 focused on the design and impact of the evaluation mechanisms towards organisational governance of the whole institutions.

Questions for stakeholders working at research institutes (or centres) at L2 focused on their responses (expectation and coordination) towards the policies of governance and evaluation made at L1, and on their interaction with L1 and L3.

How is your institute evaluation designed (purposes and criteria) and implemented?
How is institute evaluation related to institute governance concerning strategy, funding and operation?
Are the interactions with the headquarters and with researchers top-down or bottom-up?
How are the stakeholders (see Table 3.4) involved in your evaluation and governance activities? Do you have actual or potential conflicts with them?
Do you think evaluation can help to mitigate those conflicts, and how?

Table 3.7 Interview questions for L2

Questions for individual researchers at L3 focused on their personal opinions on the evaluation activities they have experienced in the framework of the institutes analysed.

What are the participatory procedures for individual researchers, before, during, and after the on-site evaluation activities?
How do you think the evaluation has an impact on individual researchers and your institute?
How do you think the evaluation criteria relates to individual researchers and your institute?
Do you think evaluation supports or hampers scientific freedom or trust, and why?

Table 3.8 Interview questions for L3

An overview of the interview data collection process is presented in Table 3.9. A pilot test with seven interviews (five on-site and two by Skype) in February 2013 helped to revise the above questions so that they could be more efficiently answered within one or two hours. This process also helped to test the first analytical framework (section 2.6). From July to November 2013, 83 individuals were contacted and eventually 57 were interviewed from the three PRIs' headquarters and 19 institutes. The purpose of the interviews was to help to answer to the research questions and to refine the analytical framework in its evolving versions. Most of the interviews were recorded and all transcribed into texts with notes for further interpretations. Importantly, detailed results of evaluation activities were all highly confidential but their general structures and impact were discussed in our interviews. The interpretation of literature and public documents was supplementary to the interview data, such as the journal interviewing the CAS president and mass media reporting the CAS reforms. Afterwards, the software tool Atlas ti, 7.0 was used to code, categorise, and integrate the dispersed and qualitative interview data. For example, many complex

discussions on organisational governance were categorised according to the three aspects (strategy, funding, operation) proposed in section 2.1.1.

Institutions	Research institutes (or centres)	Individual interviews		
		L1 Senior administration	L2 Coordinating administration	L3 Individual researchers
MPG	Headquarters	5		
	Institute A		2	1
	Other institutes			3
HGF	Headquarters	3		
	Centre B		2	4
	Centre C		3	2
	Other centres			4
CAS	Evaluation Research Centre	6	2	1
	Institute D		3	2
	Institute E		3	2
	Institute F		3	2
	Institute G		3	1
Total	57	14	21	22

Table 3.9 Overview of interviews conducted

Based on the interview data, an overview on how to analyse them to arrive at arguments that answer the research questions is explained in the next section.

3.4 Analyses Overview

This thesis chose a formative way to build the analytical instrument to guide the interpretation process of the evidence and derive some findings and conclusions. In this sense, the research started by defining a basic conceptual framework (section 2.6) that was then contrasted with evidence (literature and some interviews) and enriched to the second framework (section 4.5). The enriched framework was contrasted again empirically, leading to the final framework (section 7.1.5). This approach required multiple phases of data analyses—filling up the empty tables below step by step. First, the organisational governance of MPG, HGF, and CAS concerning the three aspects (strategy, funding, operation), as shown by Table 3.10, were discussed in Chapter 4. Descriptive analyses of the national and institutional contexts of the three PRIs provided a basis to understand their specific governance

characteristics. Along with the context analyses, the interests and power of the categorised stakeholders were researched to complete Table 3.4. The conceptual discussions on organisational governance of PRIs in section 2.1 were tested.

	MPG	HGF	CAS
Strategy			
Funding			
Operation			

Table 3.10 Governance characteristics of the three cases

Based on that, stakeholders' tensions at the three PRIs were researched considering their interests in the governance aspects and their positions and powers at the organisational levels. Particular tensions involving particular stakeholders were analysed at each level within each case study first and then compared across the three cases. Consequently, the completed version of Table 3.11—the core finding of Chapter 5—presented the similar and different conflicts at each level for the three PRIs.

	strategy	funding	operation
L1			
L2			
L3			

Table 3.11 Governance tensions at each level of the three cases

Next, the characteristics of the overarching evaluation mechanisms of the three PRIs were specified by their evaluation activities (Table 3.5). Those characteristics were individually analysed first considering whether they could help to respond to their own complex governance requirements. Then the responses regarding the three evaluation mechanisms were comparatively analysed. The qualitative comparisons were integrated in a semi-quantitative way through a spider chart in section 6.2 aiming at visualising the differences. Whether there was a common principle behind those differences was explored.

Then the balancing processes, proposed in sections 2.3 and 2.6, were analysed concerning the ways the governance tensions were addressed to and eventually mitigated by the evaluation mechanisms within the individual cases. The two hypothetical ways—institutionalisation of the evaluation mechanisms and stakeholders' interaction and negotiation—were tested. As for the institutionalisation of the three evaluation mechanisms, that is, the professionalisation and systemisation of individual evaluation activities with

considerations of and uses for organisational governance were reviewed. The idea behind doing this was to answer to the question: was there a general trend for the three evaluation mechanisms in the way they played a role with respect to organisational governance? Did they have distinct processes? These questions were answered in section 6.3.1.

The second way—stakeholders’ interactions and negotiations in evaluation processes—was traced from the descriptions of specific evaluation procedures, such as negotiating evaluation guidelines and results, across the three levels. The case studies helped to explore how stakeholders’ negotiations were designed and actually happened at different phases of evaluation activities—preparation (drafting guidelines and organising evaluators), on-site inspection (presentations and workshops), and releasing results (discussion of preliminary results and conclusions, feedback exchange, and final reporting). The phases in which stakeholders’ negotiations facilitate organisational learning were considered in the evaluation procedures. Like the comparative analyses of the governance tensions across the three PRIs, their balancing processes to mitigate the tensions were also explored comparatively.

Then, how evaluation mechanisms facilitate organisational learning was analysed. The hypothetical knowledge bases (proposed in section 2.3.1) were tested by tracing the uses and impacts of individual evaluation results and recommendations. The earlier analyses of evaluation institutionalisation and stakeholders’ negotiation were also used to support the *learning* related arguments.

The followed analyses were done to aggregate the balancing effects of the individual cases as shown in Table 3.12 (which corresponds to the filled table in section 6.5). The generalisable principles of the ideal balanced governance—mitigated conflicts—at each level were demonstrated, including both what had been achieved and what was to be pursued in the near future.

	MPG	HGF	CAS
R1 at L1			
R2 at L2			
R3 at L3			

Table 3.12 Roles of evaluation in governance at each level of the three cases

All the generalisable principles found based on the synthetic analyses of the cases are summarised in Chapter 7. The final conceptual framework (section 7.1.5) aimed to be applicable for the general PRIs beyond the three cases studied.

3.5 Validity of the Research Approach

This section discusses the validity and reliability of the mixed research approach—both quantitative and qualitative data collection and analyses of academic literature, public documents and interviews. Although replication may be not applicable for qualitative paradigms, Guba and Lincoln (1989) recommended the alternative constructs of credibility, transferability, and confirmability. Constructs of reliability of this thesis (more qualitative than quantitative) have to convince audiences that the findings are based on critical investigation, and are well founded and sound (Yin, 2009).

Originality and Reliability

This thesis provides an original integrated tool exploring co-evolution between the two key concepts of organisational governance and evaluation mechanisms. The approaches of data collection and analyses simplified very complex situations concerning the two big and abstract concepts within the large-scale umbrella PRIs. The definition and essence of the key concepts in this thesis considered academic literature, policy documents, and also the practical working languages of the PRIs. Secondary data from literature and public reports were helpful to supplement the interview data. Some joint institutes or offices and independent researchers of the three PRIs were approached to test some arguments from their employees. To enhance the construct validity, this study, on the one hand, attempted to strengthen the logical link between stages of the research design (research questions—a basic conceptual framework—individual case studies—synthetic analysis—enriched and final frameworks) (George & Bennett, 2005). On the other hand, the feasibility of each stage was justified by addressing its internal validity.

Furthermore, translating and interpreting the original data considering both languages (English and Chinese) and cultures were challenging and time-consuming but handled carefully. All research about the German cases was done in English, thanks to the large number of policies published in English and the English-speaking interviewees at MPG and HGF. Most of the original data of CAS was in Chinese, including most policy documents, part of the literature, and all interviews. Many Chinese-speaking scholars who published work in English and interacted actively worldwide were consulted to discuss our comparative analyses. Furthermore, several books related to either governance or evaluation and translated from English to Chinese were helpful to understand both technical expressions and

cultural contexts. Professional translation tools, language tutors and proof-readers all helped.

In the case studies, cultural differences in the understanding of some key concepts became visible. Notions like *conflict*, *trust*, and *balance* can have a diversity of meanings and connotations in different cultural and linguistic contexts (e.g., taken-for-granted rules) (Fukuyama, 1995) in Germany and China. This cultural factor was taken into consideration in both the discussion of the research questions with the interviewees and in the case studies. However, it is not the focus of this study.

Process of Quality Management

It has been a learning and exploratory process conducting this four-year research with formal and informal learning-by-doing. The key concepts in this thesis and their interdependences became richer and more valid in management processes from initiatives, basic outlines, planned field work, data collection and analysis, to comparisons and final argumentations. During this process, the quality and the comparability of the three cases, very abstract at the beginning, kept increasing along the way with deeper understanding of each case.

Ethical Considerations

Ethical issues were considered in data collection and the interpretation of the data. Before any interview, a general introduction and a question outline were given. During the interviews, the motivation and key concepts of this study were briefly explained, and ethical negotiation was discussed, like how to share the principles, rather than the content, of confidential documents. Agreements were made on their quotes in publications after privacy checks for which the author reserved the copyright. During the transcription, analyses and writing, all the names of institutions and interviewees were anonymised while the general introduction of the institutes and the interviewees' titles, like high-level executives, were agreed to be kept.

To sum up, Chapter 3 describes, step by step, the research approach to operationalise the theoretical discussions on the concepts in the case studies. Using the above approaches, the following Chapters—4, 5, 6—will present the empirical findings.

4 National and Institutional Contexts of the Three Cases

—Towards the Enriched Conceptual Framework

Following the above research approach, this chapter discusses the national and institutional contexts of organisational governance of Max Planck Society (MPG), Helmholtz Association (HGF) and the Chinese Academy of Sciences (CAS) and the role of related stakeholders. The descriptions and analyses of organisational governance of the three cases in this chapter provide the basis for exploring, in the following chapters, their governance tensions and how the tensions are balanced by the evaluation mechanisms put in place in those institutions.

Specifically, the national public research and development (R&D) landscapes of Germany and China and their challenges identified by the existing literature are discussed in section 4.1. The institutional contexts of the three publicly funded research institutions (PRIs) are illustrated in section 4.2. Influenced by the contexts, the governance characteristics of MPG, HGF, and CAS in terms of strategy, funding and operation are analysed in section 4.3. Based on the governance characteristics, the interests and power of various stakeholders are analysed in section 4.4. All these analyses validate the basic conceptual framework discussed in section 2.6, and leads to the identification of an enriched framework, which is explained in section 4.5.

4.1 National Public R&D Landscapes of Germany and China

This section presents the different characteristics of the national public R&D landscapes of Germany and China, and their governance challenges at the system level. Some challenges are shared by the two landscapes and have similar impact on the organisational governance of the PRIs. Based on government research reports, such as the Commission of Experts for Research and Innovation (EFI) report⁸ (2012), and discussions in international top journals, such as *Nature* and *Science*, since 1990s, it is clear that both Germany and China have embraced significant growth and specialisation in cutting-edge R&D in fields and activities with their own competitive advantages as well as challenges.

⁸ EFI (in German language: *Expertenkommission Forschung und Innovation*) was established by the German government and its reports are reliable for understanding the German R&D landscape.

Overall, this thesis concludes that Germany has a comprehensive R&D landscape and relatively clear categories of PRIs missions compared to China. As Wolf (1995, p. 832) claims, Germany's reunification and institutional reconstruction of non-university publicly funded research was rapid, profound, complete and clearly directed towards a comprehensive model in the 1990s. In the following two decades, as the EFI report (2012) shows, Germany's R&D landscape has become highly diverse with a complex funding structure. Its R&D intensity, that is, gross domestic expenditure on R&D (GERD) in relation to gross domestic production (GDP), increased continuously and reached 2.98% in 2012 at **€69.7 billion** (EFI report, 2012, p. 34). This achievement is close to the 3% target as set by the European Council in Barcelona in 2002 and on the way to the Germany's 3.5% target for the year 2020 (EFI report, 2015, p. 40). Approximately one third of the total R&D expenditure was channelled into public R&D activities, mostly in non-university PRIs (**15.8%**) and universities (**12.8%**), both of which compete and collaborate actively (EFI report, 2012, p. 56).

The federal and Länder (states) governments take different responsibilities for funding public scientific institutions, programmes, and projects in their respective spheres. Universities are mainly financed by state governments (HRK, 2015).⁹ BMBF (Federal Ministry of Education and Research), as a representative of the federal government provides the majority of the long-term R&D funding for the four non-university PRIs, namely, MPG, HGF, Fraunhofer-Gesellschaft (FhG) and Gottfried Wilhelm Leibniz Science Association (WGL). The four PRIs are relatively independent as how to spend the institutional funding they receive from BMBF and Länder governments. Still, according to the interviews, BMBF negotiates with each of the four PRIs on their funding priorities and evaluation criteria. Furthermore, the federal government legitimised the Freedom of Science Act ¹⁰ in 2012 to apply to the non-university PRIs to grant them greater financial and administrative flexibility, in terms of collaborative research projects, construction of facilities, and regulation for procurement of goods and services, to enhance the effectiveness and efficiency of the investment (EFI report, 2012, p. 54).

⁹ HRK (German Rectors' Conference) authoritatively represents the voice of the universities in Germany; website: <http://www.hrk.de/activities/higher-education-finance/>.

¹⁰ In German: *Wissenschaftsfreiheitsgesetz*, the full title 'Law on the Flexibility of Budgetary Legal Framework of Non-university Research Institutions'.

Figure 4.1 below, proposed by Heinze and Kuhlmann (2008, p. 890), reflects the comprehensive and categorised German public research system in which universities and the four non-university PRIs develop differently in light of the numbers of publications and patent applications from 1990 to 2002. Such different developments in terms of publications and patents results from the varied research orientations—fundamental, strategic and applied—categorised in this thesis. The specific research orientations and public funding of MPG and HGF will be explained in the next section. Generally, MPG focuses on fundamental research and HGF conduct strategic research to solve disciplinary and complex challenges for society. FhG and WGL are not included in this study because the former is oriented towards technological markets, which is not a focus of this study, and the latter has less clearly delineated features in relation to the governance aspects of this study. According to the annual reports, both MPG and HGF received around 5% annual increases of institutional block funding from 2011 to 2015, from both federal and state governments. Moreover, they apply for competitive project funding from many sources, including BMBF, the German Research Foundation (DFG) and also from the private sectors (*ibid*).

Figure 4.1 Institutional dynamics in the German research system between 1990 and 2002¹¹

Source: Heinze and Kuhlmann, 2008

¹¹ R&D full-time equivalent personnel are referred to by the x and y axes in their paper.

Despite a relatively stable and wealthy R&D funding system, there is an ongoing discussion on how to improve further the German R&D landscape in light of the tensions in funding allocation between universities and non-university PRIs, among various PRIs, as well as between the federal and state governments. According to an interview with the Head of BMBF in the journal *Science* in 2014, there was a debate about the allocation of €9 billion public R&D funding, which was considered as a battle about block funding between the university and non-university public institutions (Kupferschmidt & Vogel, 2014, p. 791). A suggestion was proposed to allow universities and non-university institutes to merge into cooperative units eligible for federal funding (*ibid*). Therefore, the discussions on investment in universities and various PRIs lead to a highly relevant issue—the evaluation of their relative roles and performance, given their different missions and research orientations.

The WR conducted the first systematic evaluation of all the four PRIs in Germany in 2001 with impact analysis at the national level. WR is funded by the federal and the 16 state governments and it has 32 members representing all major stakeholder organisations in Germany. It is the science policy advisory body for the structure and development of higher education and research in Germany, including evaluations of research institutes, programmes, disciplines, etc. According to Block and Krull (1990, p. 833), WR's evaluative and advisory initiative in the 1990s can be seen as a shift from input towards output assessment and an important tool for strategic change in the German public R&D landscape.

As for China, the country has been continuously boosting and increasingly developing its R&D landscape towards the leading role worldwide, a fact that is strongly supported by the literature and the interviews. As shown by *Nature Index* (2015, p. 54), *Nature Outlook* (Grayson, 2015, p. 11) *Nature* (Ni, 2015, p. 58) and *Science* (Sun & Cao, 2014, p. 1006), China's GERD increased to **€145.18 billion**¹² in 2013 as the world's second largest after the United States, and increased by an average 23% each year since 2005; The GERD/GDP increased from 0.9% in 2000 to 1.98% in 2012 and is expected to be 2.5% by 2020. As shown by *Nature* (Ni, 2015, p. 58), around 75% of total R&D investment was spent in industries that also contributed the majority of the GERD increase. About **15%** of the GERD was spent in PRIs, similar to Germany, and **7.2%** in universities, less than in Germany.

¹² 2013 conversion rate: 1 euro = 8.16 Chinese yuan.

Supported by this stably increasing R&D investment, China has been reforming the science and technology (S&T) system since 1978 increasingly to become a strong international competitor and collaborator, aiming for the proper balance between domestic efforts and engagement in the international scientific communities (AeA report, 2007, p. 121). The Medium- and Long-term Plan for S&T Development 2006–2020 (MLP) of China provides a blueprint for its transformation into a domestic innovation-driven economy by 2020 and a world leader in S&T by 2050 (*ibid*). The 12th Five-Year Plan for S&T Development (2011–2015) provides specific regulations and suggestion for implementing the MLP (*ibid*).

However, even with rich R&D investment and large-scale plans, China is ‘not getting a good return in terms of both basic-science breakthroughs and commercialisation of related findings’ (Cyranoski, 2014, p. 468). Why? This study could not answer this complex question but can suggest one reason inspired by the study of CAS—the *mixed and not well categorised research orientations* of the national key R&D contributors. This reason is intensively discussed in the literature and verified in the interviews.

The three top actors in China’s public R&D system are acknowledged by Sun and Cao (2014, p. 1006) in terms of R&D expenditures—the Ministry of Science and Technology (MOST), CAS, and the National Natural Science Foundation of China (NSFC). According to Ni (2015, p. 59), in 2013 nearly one third of the central government R&D expenditure was distributed to these three actors. As the case study shows, MOST is responsible for organising and distributing national R&D plans and tasks through large-scale research programmes with particular research themes and policy requirements like the 973¹³ and 863¹⁴ Programmes. Some of these policy-driven programmes are explicitly entrusted to ministry-owned research groups. Most of the competitive programmes are conducted by the CAS institutes and top universities. CAS receives central government revenue for intramural research and competes for extramural funding from NSFC, MOST, and other sources (Sun & Cao, 2014, p. 1007). The other PRIs, such as the Chinese Academy of Social Sciences (CASS), are considered by interviewees as not comparable to CAS in terms of scales and investment. NSFC provides funding for PRIs, universities and individual researchers in terms of peer-reviewed competitive

¹³ National Basic Research Programme

¹⁴ National High-Technology Research and Development Programme

research projects that usually have smaller scales and more general research themes than programmes of MOST.

China's public R&D landscape is considered less diverse and aggregated than that in Germany. There is lots of duplication of work found between the key actors in their different missions, research orientations and administrations. However, so far there is no systemic evaluation of the three actors to prevent such duplication. Public research funding in China is still strongly oriented towards applied and experimental R&D. Looking at the breakdown of GERD in 2013, the largest category was technology development (84.6%), compared with applied research (10.7%) and basic research (4.68%) (Ni, 2015, p. 58). Within MOST, research funding in 2013 was spent mostly in applied research (35.1%) and technology development (29.4%), rather than on basic research (15.4%) (*ibid*). MOST's 973 (basic) and 863 (technological) programmes since 1997 are considered as attempts by R&D policy-makers to distinguish and develop different research orientations. Importantly, the national R&D funding policies started to promote new clusters and networks across individual R&D conductors with the same research orientation. For CAS, its mixed and complex missions and research orientations (details in the next section) result from, and in turn worsen, the lack of coordination and collaboration among PRIs, universities and enterprises at the system level. Public policies encourage competitive project funding in NSFC to create more diverse and interdisciplinary collaboration. As Cao and Sun (2014, p. 1006) show, China has been making efforts to build a decentralised and diversified R&D funding system but lacks top-level design, unified planning, effective coordination, and transparency in budgeting and spending.

The national contexts of MPG, HGF, and CAS introduced in this section present their relative roles and challenges within the systems. This is closely related to the conflicting expectations from external and internal stakeholders of the three PRIs, to be analysed in section 5.1. Furthermore, changes of research priorities are required by national policies like Germany's Energy Transition policy (Appendix B1 box 3) and China's Innovation 2020. Those policies call for significant responses from the PRIs in various ways, like cutting off existing institutes or programmes, creating new ones, and encouraging new collaboration forms. Overall the German institutions hold tightly onto their organisational missions and governance. CAS has been undergoing policy-oriented reforms. These will be explained in the following sections.

4.2 Institutional Contexts of the Three Cases

This section presents a basic introduction of the institutional contexts of MPG, HGF, and CAS for better understanding their specific organisational governance and evaluation mechanisms that will be analysed in more details in the following sections. As introduced in Chapter 1, PRIs' institutional contexts can be very complex involving hierarchical layers, institutes at multiple locations, various research disciplines, fields and organisational cultures, etc. As the case studies show, the defined *missions* of the three PRIs determine their research orientations and specific governance arrangements in terms of strategy, funding and operation. The missions were created both by external bodies, taking into consideration the national R&D landscape, and by internal leaders, concerned with institutional development. More supplementary information on their organisational missions, structures, and evaluation practices can be found in the appendices.

MPG

As Germany's most well-known research institution with 18 Nobel Prize winners, MPG has an annual budget of about €1.5 billion and 17,284 staff (MPG 2013 annual report, pp. 92–100). It claims the organisational mission to expand the boundaries of knowledge by conducting *the best fundamental research worldwide* (MPG annual report, 2013). The 82 Max Planck Institutes (MPIs) remain highly competitive and innovative in their respective fields by attracting the best scientists and giving them the best working conditions (MPG annual report, 2013; Schruff, 2012). In principle, leading scientists at MPIs are free to conduct curiosity-oriented fundamental research in an open framework without ties to specific applications (Schruff, 2012). This is guaranteed, according to interviews, by the high levels of public block funding (85% of the total budget) and of scientific freedom of MPG scientists. Such a high level of scientific freedom follows the tradition of the person-centred principle at MPIs—called the Harnack Principle¹⁵ (MPG annual report, 2013, p. 5).

Compared with applied research, interviewees believed that the longer-term potential of fundamental research in MPG is proven by prominent awards, like the Nobel Prize. This study considers this as protected space for MPG scientists to

¹⁵ Adolf von Harnack (1851–1930) was the first president of the Kaiser Wilhelm Society, the predecessor of MPG. The Harnack Principle has been applied for nearly 100 years in MPG.

concentrate on curiosity-driven fundamental research with little concern about funding shortages or research application. As a high-level executive at MPG headquarters explained:

One MPI, for example, is working on a potential cancer therapy, which could be very breakthrough and a societal contribution. The drug has been studied for over 20 years. Nobody knows whether this will ever become a success story. Built on such basis, the high trust system is really important for scientific work for the potentially very long period that it may take to benefit society and taxpayers—over 20 or 30 years, or even longer.

Considering the above overall context, how can MPG make sure that the best fundamental research is conducted in the *person-centred MPIs*? How can such a high level of trust of scientists be guaranteed? In part, the answer could rest with the core evaluation subjects of MPG—MPIs and their directors. MPI evaluation takes place every two years—much more frequently compared with HGF and CAS. MPG claims that winning high-level scientific prizes proves the value of its evaluation as an effective and highly efficient quality assurance instrument (Schruff, 2012). How such an instrument functions will be explored in Chapter 6.

HGF

HGF is the largest PRI in Germany with an annual budget of €3.95 billion and 33,148 staff working in the 18 national research centres under its umbrella (HGF annual report, 2014, pp. 38–39). It claims the mission of *national strategic research*, aimed at solving the major and complex problems of society, science and industry by conducting high-level research in six particular fields¹⁶ (HGF 2013 annual report, p. 3). The strategic research relies on large-scale facilities and infrastructures, called *big science* by the interviewed scientists, and responds actively to national policies (Appendix B1). The ‘big research’ is considered efficiently conducted and funded by approximately 30 highly collaborative programmes across HGF centres for each five-year period. This is called *Programme-oriented Funding (PoF)*, which will be analysed in the next section.

¹⁶ Aeronautics, space and transport; earth and environment; energy; health; key technologies, and structure of matter.

In addition, HGF is responsible for planning and operating large-scale facilities and scientific infrastructure to strongly promote international visibility of Germany as an attractive centre for science and technology. HGF experienced restructuring from scattered individual national laboratories to the comprehensive Helmholtz umbrella in the early 1990s (Appendix B1). This restructuring process provides the basis to understand the tensions among different centres, as discussed in the next chapter.

According to the HGF 2013 annual report (p. 13), *research programmes* across HGF research fields, centres, and institutes are the key evaluation subjects. The scientific quality and strategic relevance of each programme, as well as the sub-topics, are evaluated and rated by international experts every five years. PoF evaluation results provide evidence for strategic adjustments in the six research fields (see footnote 15) and financial decisions. The funding decisions in relation to each field, programme, and the centres involved are suggested by the Senate Commission and the Senate, as explained in section 4.4, to the funding bodies.

CAS

CAS is a huge PRI, with more *complex organisational missions* than all the German PRIs together. There are always debates about whether it is unwieldy. CAS comprises a centralised headquarters with 6 academic divisions,¹⁷ 13 regional branches, 104 research institutes (including 3 key botanical gardens), 3 universities, and over 400 invested enterprises located throughout the country (CAS annual report, 2013, p. 7). Also, CAS has a mandate to help preparing political decisions of the State Council, the highest government agency in the People's Republic of China. Its scientific missions span wide spectrum of research orientations, including 1) scientific frontiers for technology breakthroughs and industrial development; 2) scientific challenges to improve people's welfare while addressing resources and eco-environment; 3) high-tech challenges to improve international competitiveness and national safety (CAS website, 2015). Political objectives were updated in 2014 to achieve the Four First¹⁸ in China. In addition, CAS provides over 70% of national scientific research facilities as a public S&T platform, and has

¹⁷ Divisions of Mathematics and Physics, of Chemistry, of Life Sciences and Medical Sciences, of Information Technical Sciences, and of Technological Sciences.

¹⁸ First in achieving leap-forward development of S&T, first in building high grounds for cultivating innovative talents, first in establishing a high-quality S & T national think tank, and first in building internationally leading research institutes (CAS website, 2015).

educational responsibilities in three universities. ‘There’s no organisation in the world with so many functions’ (Cyranoski, 2014, p. 468). A variety of research orientations result from decades of self-development of CAS institutes. For a very long time, there were no obvious institutional boundaries of the research orientations within CAS and even within one huge institute. How could such complex missions be evaluated as to whether or not they have been achieved?

The complexity of the CAS missions is the key to understanding its specific governance characteristics (discussed in the next section) and governance tensions (discussed in the next chapter). According to Liu et al. (2011, p. 920), CAS developed out of the planned economy approach and faced many serious challenges during the transformation towards a market economy in the 1990’s. As Luo et al. (2015, p. 61) propose, it has experienced a development from loose-knit networks towards an integrated association. As the case study will show, CAS has been making efforts over the years for strategic planning and efficient governance of fragmented research in the large number of heterogonous research institutes. Several institutional reforms, such as, Knowledge Innovation Programme (KIP), Major Outcome Orientation, and Pioneer Action Plan (PAP), aim to simplify and modernise the complex system and to align it with international norms. These reforms will be analysed in the following sections and supplementary information can be found in Appendix C3.

As shown by Luo et al. (2015, p. 63), CAS experienced a long period without evaluation and funding allocation procedures were mainly based on hierarchical decisions by the headquarters and senior professors in each field. However, R&D evaluation became a key notion in a series of institutional reforms in CAS to depoliticise and depersonalise research funding (Bai, 2007, p. 110). According to the report from CAS Evaluation Centre (2013, pp. 42–46), the current *Expert Diagnosis Assessments* for CAS institutes aims to promote their significant and promising R&D and support institutes’ long-term development. This evaluation is hereafter referred to as the *One–Three–Five evaluation* since what is evaluated for each institute is its one clear position, three breakthroughs and five key developing directions (*ibid*). *One position* of an institute means its current strengths and ranking in upper (or lower) echelons in the scientific communities, including its long-term target. *Three breakthroughs* refer to the most promising R&D targets within the institute to achieve its future position. *Five key fostering directions* are specific research topics where most of the existing resources are invested to realise

the breakthroughs. How the *One–Three–Five evaluation* responds to the changing governance, such as Major Outcome Orientation, will be analysed in Chapter 6.

The institutional contexts in this section frame both the organisational governance and the evaluation activities of MPG, HGF, and CAS. Across the three PRIs analysed, their different missions and research orientations are supposed to entail different objectives and approaches to the evaluation of research and the research units. The specific organisational governance of the three PRIs will be discussed in the next section.

4.3 Organisational Governance of the Three Institutions

As defined in section 2.1.1, the organisational governance of PRIs in this thesis involves the ways PRIs' activities and rules of science and scientists are produced, sustained, and regulated among stakeholders in relation to specific governance arrangements on aspects of strategy, funding and operation. Such processes are challenging as they concern the complexity of the national public R&D landscape (section 4.1) and institutional contexts (section 4.2) involving multiple stakeholders (summarised in section 4.4). Table 4.1 shows *a sketch of the characteristics of the organisational governance at MPG, HGF, and CAS*.

The 11 dimensions in Table 4.1, while not totally independent from each other, are summarised from the three individual case studies to extend the conceptual dimensions of organisational governance discussed in Chapter 2. The selection of the 11 dimensions responds to the definition of *organisational governance* by Hermanson and Rittenbery (2003, p. 27)—a process dealing with the procedures utilised by the representatives of the organisation's stakeholders to provide oversight of risk and control processes administered by management. Therefore, commonalities and differences of the governances of the three PRIs in terms of the selected dimensions are presented in the last three columns of Table 4.1. Some of the governance characteristics are unique but most of them are shared by two or three of the PRIs studied. The shared governance characteristics, such as institutional block funding and competitive project funding, are considered as typical elements of organisational governance of other PRIs beyond the three cases.

Based on the analyses of organisational missions and research orientations in the previous section, this section focuses on the specific governance arrangements of the three PRIs in aspects of strategy, funding, and operation. Other dimensions related to the three aspects will also be discussed.

The Balancing Role of Evaluation Mechanisms

Dimensions	Characteristics	Institutions		
		MPG	HGF	CAS
Organisational mission	To expand boundaries of knowledge	√		√
	To contribute to national development		√	√
	Talent training	√	√	√
Research orientation	Fundamental research	√		
	National strategic research		√	√
	Applied research		√	√
Funding framework	Institutional block funding	√	√	√
	Competitive third-party funding	√	√	√
Public accountability responses	Role of public representatives in decision-makers	√	√	
	Transparency of governance procedures	√	√	√
	Public education and outreach	√	√	
Legal status of institutes	Legally independent institutes develop autonomously		√	√
	Legally dependent institutes	√	√	√
Multi-stakeholders	External	√	√	√
	Internal	√	√	√
Organisational strategy	Long-term and short-term tasks under the organisational mission and research orientation	√	√	√
Management style	Top-down			√
	Bottom-up	√		
	Around core units like programmes		√	
Institutional change	Re-creation of new institutes or research programmes	√	√	√
	Reforms of governance procedures	√	√	√
Operation	Human resources (HR) like appointment of leaders	√	√	√
	Knowledge and technology production and transfer	√	√	√
	Collaboration	√	√	√

Table 4.1 Dimensions of organisational governance in the three PRIs

Source: the author, 2015

4.3.1 Strategy

The three case studies show common governance targets of MPG, HGF, and CAS—research excellence, efficient collaboration and good education, scientific synergies between disparate institutes, disciplines, and fields. To pursue these targets, the three PRIs take varying strategies.

MPG

The core strategy of MPG is recruiting the world's leading scientists in their respective disciplines to conduct *the best fundamental research*. This strategy depends on the high level of block funding and the person-centred structure in MPIs. These are the two key elements in funding and operation, in the next two sections, to follow Harnack Principle introduced in section 4.2. The scientific competitiveness and attractiveness of its 82 MPIs are centred on the world's leading researchers who are empowered to define their research subjects and construct teams and facilities. MPG can be characterised as utilising its high international reputation, an attractive block funding and high scientific autonomy to identify emerging research fields and to attract and retain the leading international scientists.

To assure scientific trust in accordance with Harnack Principle, each new appointment of an institute director entails a long-term commitment to a research subject or new field. Meanwhile, the research spectrum of MPG is continually evolving: new institutes are established to find answers to seminal, forward-looking scientific questions, while old ones can be closed when their research fields are widely established elsewhere or their research is not doing well (MPG annual report, 2013). This continuous renewal preserves the ambition for MPG to react quickly to pioneering scientific developments in focused research fields. An additional strategy of MPG is to foster creative and potential talent (Appendix A3).

HGF

As introduced in the previous section, HGF conducts *national strategic research* targeting solutions to complex societal challenges in six research fields. The associated HGF strategy aims to exploit systemic advantages by optimising internal coordination of the 18 large-scale and experienced laboratories with well-funded scientific environment and tenured scientists. R&D activities integrated in long-term strategic programmes within PoF bring together the internal resources and capabilities of various fields and centres. Externally, the HGF centres are found to actively cooperate and network with national and international scientific partners, such as universities and industry. Such collaboration mostly relies on the unique scientific infrastructures—large-scale and expensive—of the HGF centres such as supercomputers and research aircrafts. More examples on the funding and management of these infrastructures can be found in Appendix B3. The platform of

international research collaboration created by HGF is considered by interviewees to be of high quality and great reputation worldwide.

CAS

As shown in the case study, the complex missions and *mixed research fields and orientations* of CAS are undergoing a strategic process of simplification and categorisation towards a *Major Outcome Orientation*. This orientation proposed in 2011 refers to the most significant R&D achievements which have different definitions and evaluation criteria within different research fields. In fact, a series of institutional reforms have taken place in the past two decades to integrate CAS's functions and achievements from multiple institutes that have been developed diversely. First, *KIP* was considered as the forerunner of the modern disciplinary deployment and organisational structure by restructuring a large number of research institutes into prioritised fields. Second, *Innovation 2020* further improved the research quality by emphasising a more efficient deployment of resources through a portfolio of stimulating policies (CAS annual report, 2013, p. 4). Third, a structural transformation of the administrative Bureaus in the headquarters took place in 2013. Fourth, the *Three-in-One* principle strategically combines the three distinct roles—policy support, science exploration and education. The three are interrelated to clearly position the roles of each institute, to break down boundaries of units sharing similar roles, and to strengthen their collaboration (findings from interviews at headquarters). Fifth, the *Pioneer Action Plan* (PAP), starting in 2014, was reported as 'the biggest reform in its history' by the journal *Nature* (Cyranoski, 2014, p. 468). The detailed information of these reforms can be found in Appendix C3.

As presented by Poo and Wang (2014, pp. 4–5), to aggregate individual institutes to pursue the major outcomes of the whole CAS, PAP categorises all the existing R&D and the related research units into four distinct research orientations: Centres of Excellence, Innovation Academies, Big Science Facilities and Specialty Institutes (details in Appendix C3). PAP aims to aggregate R&D resources and capabilities of the various research units into four key orientations so that more significant achievements are pursued within each orientation. Each unit is supposed to be equipped with the integrated resources and talent of the entire CAS and scientists from different units but within the same orientation are greatly encouraged to collaborate. Therefore, *Major Outcome Orientation* is considered by this thesis as

CAS's breakthrough strategy to deepen and take advantage of the above institutional reforms. Many policies of CAS aim to coordinate the Major Outcome Orientation, in particular the Major Outcome oriented evaluation to be discussed in the next chapter.

4.3.2 Funding

Across the three PRIs studied, the funding frameworks have various channels—each with large public block budgets (**MPG 85%, HGF 70%, CAS 53%**), and other third-party funds discussed in section 4.1. It is considered highly important to keep institutional funding at a high level because 'without stability of funding you quickly lose the best people' (Good et al., 2015, p. 100). MPG and HGF receive a higher level of block funding than CAS and thus need to be more carefully justified to the public, which is quite an institutional need in Western countries (Whitley & Gläser, 2007) and which will be discussed in the next chapter.

MPG

Characterised by *person-centred funding*, MPIs are structured in departments each headed by a scientific director who is free to independently select and carry out their research activities in ideal working conditions. Basically, within the parameters defined by the institution's central administration in the headquarters, institutes themselves decide autonomously how and for what purposes their budget is to be deployed. Long-term funding commitments (frequently permanent) supplement the scientific autonomy and freedom of scientists. No serious funding competition between MPIs has been identified. MPIs collaborate with other PRIs and universities to apply for third party fund for less than 15% of the total funding (MPG annual report, 2013). Furthermore, there is special funding in MPG headquarters that can be applied for internal collaboration among MPIs.

HGF

The *Programme oriented Funding (PoF)* within HGF, mostly from the federal government (90%), has functioned as an innovative approach since 2001 to the funding of strategic research in Germany. The cross-field and cross-centre programmes cover the core finance of all the 18 centres and their institutes for periods of five years. Funding for individual programmes involve full costs of the participating centres including R&D and operational activities and related personnel and equipment (HGF annual report, 2013). From PoF I (approximately

2002–2007) to PoF III (approximately 2014–2018), the amount of funding and the scale of scientific facilities and personnel were almost doubled (HGF website, 2015, and interviews). Facing such large-scale and complex programmes, HGF relies on the *ex-ante* evaluation of PoF to allocate funding towards the programmes and the involved centres. Besides research performance, the most important criterion is the alignment of long-term policy goals of HGF which are indicated by Strategic Guidelines and Position Papers both organised by the HGF headquarters but also negotiated with the financing partners (such as BMBF). The detailed composition process of the PoF programmes and the impact of the two strategic documents are presented in Appendix B3. PoF evaluation will be further analysed in Chapter 6.

Around 20% of PoF is not tied to be used in specific programme activities. The benefits and justification of doing so are explained below:

Within the next five years, new ideas or finding will happen, and the director of the centre should have the freedom to arrange for changes of direction for the money or person the money should go to. That is why you must have some flexible funding. (Managing Director of HGF centre X)

In addition to PoF, the HGF centres get support in other forms. The state governments contribute to financing the centres (10% of the HGF's total budget) especially in terms of coordinating the regional construction and management of large-scale scientific infrastructure. Additionally, the amount of third-party funding in the HGF centres was increased to €1.26 billion in 2013 with an increase of 7% over the previous year. There is a special Initiative and Networking Fund of €70 million euros in the HGF headquarters to encourage various kinds of collaboration among centres in broad ways.

CAS

As the case study presents, CAS had about €4.7 billion in 2012 with a ratio of 54:46 between public block funding (both central and regional governments) and third-party competitive project funding. According to interviews, block funding of CAS is calculated according to the number of official posts, and basically covers personnel salaries and welfare. The competitive project funding is usually categorised as longitudinal and horizontal funding. Longitudinal funding is mostly from MOST and NSFC, usually in terms of large-scale and strongly policy orientated programmes. Horizontal project funding mostly comes from enterprises, via both short-term

contracts (often 3 years' duration) and entrepreneurial innovation platforms (3–5 years).

Compared with MPG and HGF, CAS has a *shortage of block funding*. Over half of the interviewed scientists agree that their fixed income barely covered the basic living expenses; about 70% of their total income is from multi-channel competitive research projects. Moreover, salaries of graduate students and many other operating expenses of scientific facilities are mostly covered by competitive project funding (interview findings). According to an interview with the CAS president Bai by Poo and Wang (2014, p. 7), complex competition for research projects and grants in CAS institutes and between scientists leads to unnecessary competition, duplication, and fragmentation of research, which will be discussed in Chapter 5.

4.3.3 Operation

According to the three case studies, the initiatives, formulation and priorities of strategic goals at L1 guide tailored operation at L2 and L3 to ensure the goals are consistently being met in an effective and efficient way. As shown by Table 4.1, to operationalise strategic targets, the three PRIs share the following actions—HR management, knowledge and technology production and transfer, collaboration, and institutional change and learning. Across the three cases, operational power at L2 depends on the *legal status and independence of research units* which is usually decided by public laws and regulations.

MPG

MPIs are legally dependent on the whole institution in terms of funding and administrations (with three exceptions¹⁹). On the one hand, the scientific development of MPIs is quite independent thanks to their *Harnack Principle*. On the other hand, the legal dependence of MPIs on headquarters makes their high level of *scientific autonomy* controllable. Some statutory rights in relation to financial and operational decisions (e.g. block funding, the signing of labour contracts) still belong to headquarters. In this way, the financial and operational autonomy of MPIs is restricted by headquarters. This governance balance at L2 between the decentralised research with high level of scientific independence and centralised operation is only observed at MPG.

¹⁹ Two MPIs are from legally independent foundation, and one MPI is an associated member of HGF (MPG annual report, 2013, pp. 88–90).

More collaboration with universities is found in MPG and HGF than in CAS, since CAS has internal universities with the right to award education degrees (like doctorates), while in Germany such a right belongs exclusively to universities. According to their annual reports, MPG has been establishing new institutes preferably nearby universities for institutional connections. The establishment of International Max Planck Research Schools (IMPRS) at universities further strengthen the existing networking between MPG and universities. The IMPRS, whose leaders are usually MPI directors, should take both *ex-ante* evaluation before being granted for six years and *ex-post* evaluation in the fifth year to decide whether it can be prolonged. The considerable international spread of IMPRS promotes the reputation of MPG and its attraction of excellent scientists.

HGF

As discussed, HGF targets systematic governance of their 18 centres by optimising their *scientific coordination* through integrated research programmes. However, the HGF 2013 annual report indicates that individual centres still take key responsibilities for scientific competence and infrastructures. According to interviews, as long the strategic guidance is followed, the legally independent centres have the autonomy to define and manage the programmes based on their own circumstances. The administrations of the interviewed centres, which have secondary institutes working on sub-topics of the centres' programmes, claim to operate independently. Larger-scale centres usually operate in more than one programme and collaborate with multiple centres (HGF annual report, 2013, pp. 16–30). Furthermore, with the use of 20% of their funds not fixed by PoF, the interviewed centres agree that there is sufficient flexibility to operate additional research activities beyond PoF proposals.

Networking with universities is also important for HGF centres in terms of collaborative programmes, teaching practices and joint professorships. Each HGF centre has its own graduate school responsible for structuring PhD education and annual evaluation through hard skills of scientific qualification, and soft skills of tutoring students and communication.

CAS

As the case study shows, the powerful CAS headquarters with centralised resources takes charge of the *overarching operation of CAS institutes*. The headquarters govern the operation of CAS institutes through budget freezes and restricting the

quotes of CAS internal project application and of recruitment. Strictly following the headquarters' rules, directors have the responsibility to operate CAS institutes. Among the various units at L2, some institutes are legally independent, while some are not, due to complex reasons—history, geography, number of personnel, etc. According to interviews, it matters significantly whether one institute is legally independent or not. For example, one interviewee explained:

If you are independent, you have much more right to develop by yourself, getting more funding, projects, attracting more talent, making your own financial and personnel regulations. If not, you have to follow headquarters rules strictly and rely on the affiliated centre in terms of the limit numbers of programme application, tenured staff, students, etc.
(Director of R&D Management Department of CAS institute d)

The CAS reforms discussed in section 4.3.1 are considered helpful to improve the operation of CAS institutes towards concentrated research orientation, simplified administration and autonomous development. Particularly, Major Outcome Orientation is considered by the interviewees as a sign of increased scientific empowerment from L1 to L2, as long as the significant outputs of CAS institutes are still under the control of the headquarters. More operational autonomy is reflected by more open policies made at L1 to promote diverse international collaboration, translation and transfer of knowledge and technology at L2. The policy examples can be found in Appendix C2.

The analyses of the organisational governance of MPG, HGF, and CAS in this section validated the three aspects of PRIs' organisational governance (strategy, funding, operation) and the related conceptual discussions in Chapter 2. Concerning the above governance aspects, key stakeholders involved in the organisational governance of the three PRIs will be discussed in next section.

4.4 Categories of Stakeholders of MPG, HGF, and CAS

The key stakeholders involved in the organisational governance of MPG, HGF, and CAS are categorised in Table 4.2, which includes a summary of their composition and concerns. The three PRIs share more categories of external stakeholders than internal decision-makers such as B④ (international scientific communities). Most of the categories of *internal stakeholders* are shared by the three PRIs. Detailed positions of the internal stakeholders (category A) can be found in the sections of *organisational structure* in the appendices.

The Balancing Role of Evaluation Mechanisms

Who are they?		How are they constituted and what are their concerns?
Internal stakeholders A	① Senate and Senate Committee (MPG and HGF) Leading Party Group in CAS	The highest decision-makers. In MPG and HGF, in addition to Scientific Members, the externally appointed members include representatives from governments, academic life, business, and industry. In CAS, internal members of Leading Party Group (which refers to the Communist Party of China), without external representatives.
	② Executive Committee Leader Group in CAS	Presidents, vice presidents, scientific or administrative directors of all the institutions. They are in overall charge of research policies of all the institutions supported by the headquarters.
	③ Scientific Members (CAS calls them Academicians)	The scientific directors of centres or institutes. They contribute to decisions of all the institutions and of research centres and institutes.
	④ Headquarters	Departments, divisions, Bureaus supporting Executive Committee in institutions' administrations.
	⑤ Scientific Advisory Board	Internal or external experts for scientific consultancy. They contribute to scientific decision-making and evaluation.
	⑥ Member centres and institutes and their leaders	The unit responsible for R&D production. Boards of directors are in charge of research centres and institutes.
	⑦ Individual researchers	Senior and junior researchers participating in R&D.
External stakeholders B	① Central (federal) and state (Länder) governments	The funding bodies of PRI who make science policy in national systems. Representatives in the Senate Committee in MPG and HGF.
	② Research Funding Council	DFG in Germany, European Commission, NSFC in China, all as third-party funders.
	③ Advisory Councils	WR provides advice to the federal government on the structure and development of research; some members are MPG or HGF scientists. China Association for Science and Technology is the largest national non-governmental organisation of scientific and technological workers; some members are CAS scientists.
	④ Scientific community	International scientific institutions and experts who are often evaluators.
	⑤ Universities	Research collaboration in joint professorship, training researchers, and application for third-party funding.
	⑥ Industrial enterprises	Collaborators in R&D and technology transfer.
	⑦ The public	The general public and mass media concerning R&D.

Table 4.2 The categories of stakeholders of the three PRIs

Source: the author, 2013

This thesis compares the different interests and powers of the categories of stakeholders of MPG, HGF, and CAS and particularly their tensions on organisational governance. This section will not discuss all items in Table 4.2 in a repetitive manner but provides the basis for exploring the specific stakeholders' tensions concerning organisational governance at MPG, HGF, and CAS in the next chapter. For example, there is a significant difference shown by A① in Table 4.2 on the composition of the highest decision-making and supervisory bodies—Senate in MPG and HGF and Leading Party Group in CAS. There are external senators in MPG and HGF from various backgrounds: science, business, politics, the media, and other major sectors (MPG and HGF websites, 2013). Such composition is considered by interviewees to involve expertise and experiences of key internal and external stakeholders in Senate decisions. The decisions in MPG focus on the establishment or closure of MPIs. HGF's Senate deliberates on decisions of thematic and funding priorities for research programmes. The Leading Party Group in CAS overlaps with most of the members of its Leader Group and neither involve any external member.²⁰ However, for all the three PRIs, *headquarters* is used in this thesis to refer to the comprehensive concept of highest *decision-making and supervisory bodies* at L1. Elsewhere in this thesis, where **headquarters** is used, involves the top four internal stakeholders in Table 4.2 (A①②③④) so as not to repeat their different titles and compositions in MPG, HGF, and CAS.

As discussed in Chapter 2, tensions of organisational governance at PRIs result from the perspectives of stakeholders—their multiple responsibilities and contributions, interests and powers. Sections 4.1 to 4.4 provided the basis to validate the sources of governance tensions, as presented in proposition P1 in section 2.5—tensions of organisational governance in PRIs relate to the three governance aspects (strategy, funding, operation) and to various perspectives of the involved stakeholders on the three aspects that are embedded in national and institutional contexts. Therefore, to answer the overarching research question—how do evaluation mechanisms of PRIs play a balancing role and help to mitigate tensions in organisational governance—the sub-questions and propositions should be updated according to

²⁰ There is only one external member, working inside CAS, representing the Central Commission for Discipline Inspection of the Communist Party of China.

the achieved empirical findings. Moreover, the basic conceptual framework proposed in section 2.6 can be enriched, which will be discussed in the next section.

4.5 The Enriched Conceptual Framework on the Role of Evaluation Mechanisms in Organisational Governance

The findings in sections 4.1 to 4.4 help to enrich the basic framework presented in section 2.6 by elaborating the key concept of *organisational governance* in this thesis. Various stakeholders in national and institutional contexts proved to be influential in relation to PRIs' organisational governance in terms of strategy, funding and operation. The internal stakeholders identified validate the multi-level governance framework proposed in section 2.1—central decision-makers and supervisory bodies at the macro level (L1)—who make internal decisions and interact with external stakeholders; research institutes at the meso level (L2) under the organisational umbrella that comply with overarching policies but conduct complex R&D activities; and individual researchers at the micro level (L3), who participate in specific governance arrangements. Based on the three case studies, Figure 4.2 captures the overview of involvement and interaction of the key stakeholders across the three levels with their hypothetical controversies (bubbles: C1, C2, C3).

Figure 4.2 Stakeholders' involvement and interaction concerning governance conflicts

Source: Luo et al., 2015²¹

At the macro level (L1), the governance concerning strategy and funding is found to be influenced by the government funders and relative roles and performance of competitors, such as the competition for public funding between PRIs and universities discussed in section 4.1. The requirement for the justification of public funding vis-à-vis external stakeholders, discussed in section 2.2.1, is proved strong in the three cases. Furthermore, section 4.2 showed that decision-makers at L1 justify public funding by achieving the claimed organisational missions related to different research orientations (fundamental, strategic, applied, or mixed). Thereby, different subjects and approaches to evaluation are used by the three PRIs to measure whether their missions are achieved. It is a challenge for them to make sure that the evaluations procedures and results are convincing for the multiple stakeholders.

Additionally, some governance targets discussed in section 2.1 are validated, such as ‘better strategic alignment’ from L1 to L2 as Cozzens (2007, p. 238–239) calls it, or, as Gläser (2007, pp. 256–257) points out, ‘the aggregate effects of the individual responses towards research assessment system’ at L2. How these targets, expected by L1, are interpreted and responded to by various stakeholders at L2 will be explored in the next chapter.

At the meso level (L2), research institutes interact with multiple stakeholders from both L1 and L3 on various aspects of governance. The stakeholders’ tensions supposedly result from their diverging interests and powers, such as resource allocation among institutes at L2. As section 4.3.3 shows, this hypothetical tension relates closely to the degrees of scientific autonomy and legal independence of institutes. Based on the work of Benz (2007) and Heinze and Kuhlmann (2007), various institutes conducting research in different fields are found to need a balance between cooperation and competition with other partner institutes under the same umbrella and with external research institutes. Additionally, there is another typical mediation situation at L2 (March, 1991; Baumann, 2015)—the choices between exploration of new research directions, institutes, programmes, and exploitation of the current resources. All the above hypothetical tensions are explored in the case studies of MPG, HGF, and CAS. The comparative findings will be presented in the next chapter.

²¹ Figure 4.2 and the relevant discussions are published in *Asian Research Policy* 6(2015).

At the micro level (L3), a key conflict has been discussed by scholars, that is, a conflict in the relationship of scientists' trust and freedom within PRIs. Scientists' freedom and trust are of great importance in science because 'the reliability of scientific testimony ultimately depends on the reliability of the testifier; an untrusting, suspicious attitude would impede the growth of knowledge, perhaps without even substantially reducing the risk of unreliable testimony' (Hardwig, 1991, p. 707). Based on the work of Cruz–Castro and Sanz–Menendez (2007, p. 209) and Whitley and Gläser (2007, pp. 10–11), this thesis assumes that a higher proportion of block funding and extra flexible funding provide researchers with more trust and freedom about the selection of research topics and approaches, and more time to pursue ambitious intellectual goals. According to Hansson (2006, p. 162), research management concepts have become more oriented towards bottom-up and empowerment of organisational trust in researchers. Whether the goals of higher block funding and more trust for scientists conflict with other stakeholders is to be tested.

Consequently, the original sub-questions and propositions on the sources of governance tensions, proposed in section 2.5, get validated. To answer the overarching research question—how do evaluation mechanisms of PRIs play a balancing role and help to mitigate tensions in organisational governance—a **new set of sub-questions and propositions** are proposed. These new ones become more explicit to guide the empirical analyses of the three PRIs in the following chapters.

- ① What are the typical governance tensions of the three PRIs concerning both governance aspects (strategy, funding, operation) and multilevel stakeholders (L1, L2, L3)?
- ② How do the evaluation mechanisms of the three cases reflect and mitigate governance tensions concerning the sources identified—governance aspects and stakeholders?
- ③ What are the balanced organisational governances for the three cases?
- ④ Do the balancing roles of evaluation mechanisms of the three PRIs perform differently in their different contexts, and if so, is there a generalisable principle?

The following are updated propositions:

P1: The different governance tensions of MPG, HGF, and CAS could be categorised, both aspect-wise and level-wise.

P2: Particular governance conflicts are targeted by the relevant evaluation activities, which impact the design and institutionalisation of the overarching evaluation mechanisms.

P3: The balanced governance involves fewer tensions from multiple stakeholders across the three organisational levels.

P4: There should be a generalisable principle to be applied generally for PRIs beyond the three cases.

Consequently, the basic conceptual framework is enriched by several new elements as illustrated in Figure 4.3. The wider frame of national contexts and the narrower frame of institutional contexts are confirmed to embed both organisational governance and evaluation mechanisms. The key external stakeholders—governments and others (research funding council, scientific communities, collaborating universities, and industries presented in section 4.4)—are positioned into the national contexts. The three-level pyramid organisational structure is validated and added into Figure 4.3 that involves key internal stakeholders—decision-makers at L1, research institutes at L2 and scientists at L3. This structure is also assumed to work for evaluation mechanisms to be tested in Chapter 6.

The core element—governance conflicts—the black box in the centre of the framework (Figure 4.3) is proved both aspect-wise (strategy, funding, operation) and level-wise (L1, L2, L3). Next, the black box will be elaborated in the next chapter by analysing its specific performance on strategy, funding and operation and involving multi-level stakeholders within particular contexts of the three PRIs. As discussed, evaluations at MPG, HGF, and CAS are aligned with governance uses. How? The mutual impact between organisational governance and evaluation mechanisms will be ascertained by empirically analysing how governance tensions are responded to by evaluation mechanisms.

Figure 4.3 The enriched conceptual framework on the role of evaluation mechanism in organisational governance

Source: the author, 2015

To sum up, this chapter introduced the national and institutional contexts of MPG, HGF, and CAS to analyse their organisational governance and the involving stakeholders. In their given national contexts the three PRIs have adopted different roles and show different performance. In their institutional contexts, MPG, HGF, and CAS aim at different missions by conducting different types of research. These missions and research orientations guide their governance rules and arrangements concerning strategies, funding, and operation. Overall, the three cases share typical elements of organisational governance as shown in Table 4.1; they also share categories of key internal and external stakeholders as presented in Table 4.2. Findings on their governance characteristics in terms of strategy, funding, and operation and the involved stakeholders help to validate the sources of governance tensions. Therefore, a new set of sub-questions and propositions are updated, which enriches the conceptual framework shown by Figure 4.3. The next two chapters will explore the specific governance tensions and evaluation mechanisms at MPG, HGF, and CAS as well as the relationship between the two concepts. New elements will be further added for the final comprehensive conceptual framework of this thesis.

5 Governance Tensions at MPG, HGF, and CAS

The detailed analyses of the contexts and stakeholders in Chapter 4 provide the basis to understand the main findings of this chapter—governance tensions at Max Planck Society (MPG), Helmholtz Association (HGF), and the Chinese Academy of Sciences (CAS). The enriched conceptual framework (Figure 4.3) is used to identify specific governance conflicts that are summarised in Table 5.1. Based on this table, a comparison of the formation and impacts of specific tensions involving the multi-level stakeholders is presented in this chapter. Sections 5.1 to 5.4 discuss how the governance tensions come about between particular organisational levels involving key stakeholders of the three PRIs. Section 5.5 summarises the types of conflicts identified in terms of strategy, funding, and operation.

Regarding the concept of *governance tensions*, it is worth mentioning that the German interviewees showed a more open and neutral attitude towards such concept, as well as its synonyms (conflicts and controversies). In contrast, the Chinese interviewees showed a more sceptical attitude. Yet, all the interviewees preferred to use the word ‘difference’ or ‘divergence’ instead of ‘tensions’. Conflicts on solid issues (infrastructures and incentives) were discussed more among the Chinese interviewees. German interviewees talked about conflicts in more abstract ways, like thinking and cooperating in communication and negotiation between stakeholders and groups. However, this chapter focuses on presenting the empirical results on *governance tensions* of the three cases, but does not explore the complexity of cultural understandings and interpretations of *governance tensions*.

Table 5.1 summarises the governance tensions related to MPG, HGF, and CAS in the three-by-three matrix proposed in section 3.4. Table 5.1 is constituted by governance aspects (C1, C2, C3) and organisational levels (L1, L2, L3). Based on the cases observed, each conflict happens between particular stakeholders under particular contexts in each case. Such a three-by-three matrix summarising the tensions seek to include all the observed facts and patterns of tensions, both actual and potential tensions. Some divergence has existed for a very long time and some is only developing. These different development phases will be analysed below when illustrating specific tensions. Overall, this table helps to answer the question—what needs to be balanced at the three PRIs? The specific examples and

evidence of the tensions involving multi-level stakeholders are illustrated level-by-level in sections 5.1 to 5.4.

Organisational levels	Governance conflicts		
	C1: Strategy	C2: Funding	C3: Operation
Macro level L1	External justification of public funding versus internal governance of heterogeneous research	Deciding resource allocation principles that are impacted by many factors, such as scientific performance	Overarching policies are not one-size-fits-all
Meso level L2	Conflicting goals from interpretation of overarching organisational strategies by member institutes		Autonomous power of institutes varies
Micro level L3		Competitive grants distract scientists	Scientific freedom is threatened by management instructions

Table 5.1 A summary of governance conflicts at MPG, HGF, and CAS

Source: the author, 2013

5.1 Tensions at L1 between Internal and External Stakeholders

As analysed in Chapter 2, external justification of public funding for PRIs is required by national funding bodies that are the key external stakeholders who have a strong motivation to impose standards and regulations to check the performances of PRIs. Section 4.1 presented the literature on controversial debates within the national R&D landscapes of the three PRIs about their relative positions compared to those of their collaborators and competitors. According to that section, the central government's intention—in both Germany and China—is to set priorities of science policy that guide decisions on whether and how to continue funding the three investigated PRIs. Across the three PRIs studied, MPG (85%) and HGF (70%) have a higher proportion of block funding than CAS (53%) and more collaboration with domestic universities in terms of training and short-term research projects. The case studies show that CAS faces stronger and more complex competition with domestic top universities than MPG and HGF, in terms of competitive grants from

public research foundation, e.g. NSFC in China and DFG in Germany, as introduced in section 4.1.

The three PRIs analysed have their own approaches to legitimise public funding. As presented in section 4.3.1, MPG focuses on its strictly defined scientific missions with few policy requirements. In contrast, there are much more top-down requirements by the funding bodies for HGF and CAS, in terms of strict and detailed guidelines on R&D themes and applications. According to the annual reports of the three PRIs, there are common overarching goals when the three PRIs legitimise public funding—knowledge and technology creation and transfer, training, employment, public education, and outreach, etc. All the three PRIs claim that they are on the right course to maintain contributions to their respective national and international R&D landscapes. However, the case studies show that in all three institutions, because of the authority and complexity of R&D activities, the decision-makers at L1 face the challenge of justifying the heterogeneous and confidential analyses of scientific inputs and outputs, in a transparent and easily understandable way for external stakeholders, like politicians.

This observation is more obvious in MPG because its fundamental research has fewer immediate societal applications and responds less actively to national policies compared with HGF and CAS. Moreover, the interviewed scientists always prioritise research quality and reputation in particular scientific communities over the ‘standardised’ and ‘less scientific’ expectations of the external stakeholders, especially funding bodies and the public.

On the contrary, HGF has been addressing complex societal challenges by its interdisciplinary R&D. To react swiftly to changing policies and emerging challenges, it has drawn upon diverse teams working together in an efficient manner. This is considered by this thesis as a potential conflict between policy-based regulatory science and open-ended academic science, as discussed conceptually in section 2.2.1. The societal impact of the HGF R&D is the most explicit of the three cases. That is why HGF gets the most detailed guidelines on research themes in strictly defined research fields. According to an interview with a high-ranking executive in HGF headquarters, such regulatory science with strong strategic requirements can give rise to many other external expectations such as political advice on education and internationalisation, which is far less interesting for internal stakeholders, especially natural scientists, as this interviewee posited:

Political requirements generally draw on strategic issues, with few scientific details on which scientists usually focus more. So we have to find the balance by looking broadly at our landscape over the research fields, broadly enough that you can see the entire picture of HGF and even beyond.

Additionally, some HGF centres face different political concerns from the federal and state governments when strategic programmes do not fit the local development. Both levels of government care about significance and attractiveness in science, education, recruitment, and political advice, but at the levels of the nation and regions respectively. Focused research fields in Germany attract worldwide scientists, which contributes to regional development in terms of research, politics, and mass media. Operational issues also differ a lot for the HGF centres located in big international cities or isolated regions. For example, as one of the interviewees commented:

The international standing and scientific background of HGF is positive for the development of ...[their particular city] as an international city with four universities and over 40,000 students. It is pretty clear that without the scientific impact, this city wouldn't be able to develop so fast now. Local government cares about the number of employees and wants to keep scientific impact locally to develop this city and compete with other national lighthouses. It happened that central government recommended closing down or moving some facilities that did not meet the interests of local government. (Managing Director of HGF centre Y)

More involvement and active participation of external stakeholders in internal decision-making processes is found in the two German cases through the composition of their Senate members, who come from a wide spectrum of backgrounds. Their Senate members bring about strategic interactions from diverse perspectives: politics, academics, business, and industry. However, CAS is given politically defined R&D tasks from its one and only leader—the State Council. Such a centrally planned system sets up research scenarios for CAS characterised by a close entanglement of science and politics. The other external stakeholders, such as project funders or collaborators, have very little impact on the organisational governance of CAS in terms of strategy, funding, and operation (interview findings).

The conflict between policy-based regulatory science and open-ended academic science is also found in CAS. Over the past two decades, its overarching policies focused more on national demands than institutional development. A typical example is the overarching requirement of economic impact from R&D on China's innovation system since 1990s. Especially after the economic crisis in 2008, many research programmes have been invested in, concerning greatly the capability of technology transfer of CAS institutes. Since then, innovative and applicable technologies have been pursued following a national request of all CAS institutes including those doing basic research. The following are some interviewed responses at L2 about such an overarching requirement.

Our research needs to consider seriously economic and societal impacts. At best our research can result in products into markets. This is the requirement from central government, which has the most political power. Given the practical contexts of industries and businesses, we need companies to do the D part after we have done the R part. (Director of Department of Technology Transfer of CAS Institute a)

Various doubts arose about the overarching requirement on technology transfer:

Technological research is greatly increasing due to funding policy encouragement, but it really should depend on your own position and strengths. Policy is changing all the time, but you have to stick to your own orientation and uniqueness and expand them. (Director of CAS Institute c)

The ability of the whole CAS to handle science and technology (S&T) investment risks was weak and even the attitudes towards risks were negative and conservative. They only dare to invest in those promising fields, which are usually short-term, but ignore long-term potentially important but risky fields. (High-level executive of CAS Institute b)

Meanwhile, most Chinese enterprises cooperating with CAS are not very capable of taking R&D risks and expect quick returns on short-term investment. As for some inevitable longer-term risks, for example natural-resource-related R&D, if enterprises are not able to or would like not to take the risks, or if the market is very negative,

Ministry of S&T and some state-owned companies will take them with national financial support. (Director of Department of Technology Transfer of CAS Institute a)

The case studies show that the response of CAS as a whole towards national expectations does not fully consider internal scientific responses. Technological markets change fast so CAS's policies keep changing accordingly. This is found to be inconsistent with CAS strategies that are supposed to follow scientific development in the long term. Many institutes conducting fundamental research were found to be stressed by a heavy burden of technological requirements and forced to expand project-driven branches, which distracted many resources from their main research orientation. Such conflict between policy-based regulatory science and open-ended academic science in CAS is also related to the complex research orientations of CAS institutes driven by various short-term policies. This will be discussed in the next section.

5.2 Tensions between Headquarters at L1 and Institutes at L2

The three case studies show the tensions from L1 to L2 are mostly between headquarters' overarching governance decision-making (strategy, funding, operation) and institutes' autonomous development. Across the three cases, overall, governance decisions at L1 rely on standardised measurement of the current and future, sometimes controversial, performance and productivity of units at L2. Such overarching standardisation, for the purpose of convincing evidence, has considered adjustment according to heterogeneous characteristics of R&D in different disciplines, fields, and institutes. Yet, such adjustment has been widely discussed and debated with conflicting concerns by institutes as well as other units (centres, programmes, laboratories, etc.) at L2.

There is an imbalance identified between MPG's core strategy and the MPIs' scientific work by the case study. The key strategy—attracting the best scientists in the fields to pursue the best fundamental research worldwide is embraced by all MPIs. This long-term and stable strategy of MPG claims to be sustained by following the principle to start new promising institutes and to close down existing institutes that are underperforming (Schruff, 2012). However, as interviews show, in fact, many new institutes have been established but very few institutes have

been closed.²² Why does this happen? Interviewees explain that the potential and protected space for outstanding scientists in MPG to obtain top awards for fundamental research findings necessitates tolerance of scientific risks in very long term. This reflects the high level of trust towards MPG scientists. The interviewed decision-makers at L1 claim that they still follow the principle of possibly closing down that ongoing research of existing institutes if those institutes are underperforming. Based on that, the related challenge—to assess professionally the current performance and long-term potential of existing institutes—has been entrusted to evaluators, which will be discussed in the next chapter.

Conversely, for units at L2 in both HGF and CAS, the tension between organisational strategy alignment and autonomy for legally independent centres or institutes is identified. As introduced in Chapter 4, most of the oldest HGF centres, with well-off facilities and experienced scientists, have been powerful for decades, even before the restructure of the HGF umbrella organisation. The autonomous HGF centres are capable of developing R&D activities beyond national policy guidance and expectation. An example was given by one interviewed scientist:

Under the general mission of HGF, our institutes have their own specific work. For example, our strategic profile is to study a kind of disease that scientifically includes type I and type II. Within each type, we look at aspect 1 and aspect 2, ranging from very basic to applied research. We have several programmes cross these aspects with partners inside or even outside HGF. The HGF mission could not supervise each scientific aspect, not to mention those external aspects. We define and do them by ourselves. (A scientist from HGF Centre X)

The HGF headquarters does not have much power to control such tension because:

The centres have evolved and have a lot of resources already. If their performance is not good enough, what can we do? Do we really have the force or a tool to make them change? What we do is to fix slightly the budget before each five-year period, not to distribute money explicitly for research. We do not close down any research in any centre, just guide and recommend their directions. (High-level executive from HGF headquarters)

²² A few MPIs were restructured or renamed.

The similar tension between the headquarters' overarching strategy-making and institutes' autonomous development is also found in CAS. The large number of units at L2—research institutes, universities, laboratories, technology support centres, companies, etc.—and their diverse development make the unified organisational governance impossible. To aggregate R&D capacity of its 104 research institutes at L2, CAS has made many reforms in the past two decades like KIP (Knowledge Innovation Programme), Major Outcome Orientation and PAP (Pioneer Action Plan), introduced in section 4.3.1. The new governance of the 104 research institutes aims for long-term R&D development by aggregating their existing diverse and scattered research projects. According to interviews, to follow such large-scale and costly reform, various CAS institutes have to re-organise their existing mixed research; headquarters has to cope with the different responses from legally independent institutes that have been empowered to develop autonomously over the years. The effects of the reforms are not measured by this study, but the interviews show a concern that the long-lasting mixed research of CAS institutes at L2 are not fully ready to be aggregated by policy-makers at L1.

The above contradiction on the aspect of strategy in CAS is intensified by the large number of complex research project funding at L2, as introduced in section 4.3.2. The competitive programmes and projects from diverse funding sources are considered important by those interviewed at CAS institutes to make up for the unsatisfactory proportion of block funding. Yet, multiple budgetary channels mean varying requirements from different funding bodies, leading to fragmented research and instability, unpredictability, and discontinuity of R&D governance (Poo & Wang, 2014). Some of the CAS institutes where interviews took place had been very proud to develop comprehensively until they realised the challenge of increasing resources being fragmented. For example, a director of a CAS institute commented:

It is only one third of stable institutional fund that we can treat as a small stipend within the entire institute's budget. We have to rely on competitive research programmes and projects although we know they make the whole research profile scattered and short-term. (Director of Department of R&D Management of CAS institute c)

This financial situation is considered by this thesis to worsen the already complex and scattered research orientations of CAS at L1. It is also negative for institutes at

L2 to align with organisational strategies of CAS in the long term. Based on that, the severe financial competition between units at L2 is explained in next section.

5.3 Tensions among Institutes at L2

As shown by Table 5.1, competition for resources takes place among institutes, as well as among other units (centres, programmes, laboratories, etc.) at L2. Such horizontal conflicts are identified in the three PRIs analysed, but in different aspects and to different degrees.

Across the three PRIs studied, there was little financial competition observed between fields or MPIs that develop independently,²³ because each director gets their own fixed resources and powers. As the case study shows, because of the high proportion of institutional block funding (85%), research leaders in MPIs are not strongly motivated to undertake many competitive research projects. Meanwhile, the growing interdisciplinary character of the fundamental research in MPG gives rise to more mixed goals that extend beyond individual institutes' walls (Schruff, 2012). The independent research institutes are encouraged to increasingly collaborate on emerging mixed research goals. However, such internal disciplinary collaboration across MPIs is considered more difficult and less active than external collaboration in the same research field (interview findings). A potential reason could be the heterogeneity of their various fundamental research fields.

Interviewees at HGF believe that the comparability between the six fields is nearly zero, but programmes within the same field compete strongly with each other. The participatory HGF centres within the same programme both collaborate and compete. The tension among HGF centres between their Programme oriented Funding (PoF) competition and research programme cooperation is noted by the interviewees, especially between large- and relatively small-scale centres, and between senior and newly established centres. In terms of composing research programmes, larger centres with multiple programmes are found to have more capability and flexibility, than smaller centres with only one programme, to adjust their competition and cooperation with other centres. One example is from an interview with the Managing Director of HGF centre Y:

²³ There is an exception in international MPIs presented in Appendix A4 by the example of a partner institute of both MPG and CAS.

We work on seven programmes in one field and three programmes in the other field. There are more choices about collaboration with other HGF centres, or industries and universities. A PoF evaluation once suggested us to give up one part of research. Afterwards it was no longer funded by the HGF programme. But we still have the choice—either getting rid of it or sticking to it with our good reasons and own funding. (Managing Director of HGF centre Y)

The smaller-scale HGF centres with a limited number of programmes depend more on PoF decisions about the centres' resources available for five-year period, as one interviewee from HGF Alliance Office highlighted:

I guess there were 40% of the centres against PoF, in mind not in real actions, and 60% supportive, because the opportunity seemed less attractive for big centres with abundant resources than for small centres. Bigger centres may face emerging restrictions from HGF but smaller centres find a totally new development platform.

Therefore, each round of PoF results should be an internal balance between the 30 programmes across the six fields and, more importantly, between the 18 centres involved in the programmes. Those interviewed at HGF headquarters believe that high-quality research will result from the competition among programmes and their hundreds of sub-topics. However, as the Council of Science and Humanities (WR) evaluation report shows (2001, pp. 7–8), there was an untapped potential in the heterogeneous programmes of HGF primarily because of the unbalanced competition and coordination at both programme and centre levels. This thesis found that the problem of unbalanced research funding priorities, that emerged when the HGF brand was established in the 1990s, still remains.

As shown by the three case studies, block funding as stable long-term public support (usually no expiration) is embraced at L2 but is usually squeezed at L1 by certain rules. Unique to CAS, decision-makers at L1 actively encourage institutes at L2 to apply for short-term projects from various channels, often with duration of 3–5 years and exhausting procedures, to demonstrate their scientific competitiveness and to supplement institutional funding. The competition for diversely funded research projects is considered by interviewees to be severe and complex among heterogeneous CAS institutes and even among groups within one institute.

However, some of competitions are considered as a duplication of work and a waste of resources because the applicant institutes should have collaborated.

A key reason for the mixed and unstructured research orientations (fundamental, strategic, applied) of CAS institutes is the lack of negotiation and collaboration between them. This argument was confirmed by the case study based on the interview with the CAS president by Poo and Wang (2014, pp. 6–7). According to that interview, the CAS institutes were relatively isolated and lacked cooperation and communication because of cross-institutional barrier or geographic separation; There are nearly 20 institutes in biological sciences and more than 20 institutes in Light Emitting Diode (LED) research, all isolated from each other, all loaded with fragmented and low-quality research programmes (*ibid*). The factor of geographic separation is also verified by the regional analysis of the responses of various CAS institutes towards the KIP reform (Zhang et al., 2011, p. 875). The interviews confirm that it is not uncommon that researchers across different institutes working on related projects barely know each other. For one large-scale and complex research institute, the problem of duplicating work and limited knowledge-sharing exist among its internal groups.

Teamwork becomes the key notion in the governance reforms of CAS. As the interviews suggest, the balance moving from financial competition towards research cooperation for the 104 institutes is targeted by CAS's new strategy (the PAP reform in section 4.3.1) through the newly categorised four research clusters. That reform aims to create an environment where scientists within each cluster collaborate actively. Furthermore, a higher level of block funding and additional salary bonuses become incentives for new research clusters to relieve the funding competition between institutes, groups and scientists. A series of corresponding policies, such as credit allocation of co-authorship in publications and of joint intellectual properties, are emerging to facilitate research collaboration between CAS institutes. Not surprisingly, it would entail a long exploratory way for CAS to pursue the expectation of the reforms. The tensions at L2 in this section influence the tensions between L2 and L3 explained in next section.

5.4 Tensions between Institutes at L2 and Scientists at L3

Across the three PRIs studied, the following two tensions between L2 and L3 are widely discussed. The first is scientists being distracted by the need to apply for

competitive research grants to supplement institutional funding; the second is whether their scientific freedom, such as the ability for individual researchers to pursue their own initiatives, is threatened by management instructions and research evaluations. As observed in the three cases, depending on their scientific achievement and networking, individual scientists have autonomy in operational decisions and in interpreting organisational strategies into scientific details. A consensus emerged from the interviews at L3 that research has to be assessed but scientists' freedom has to be protected in order not to hamper the continuity of scientific research. Overall, all the interviewed researchers appreciate independence, freedom and trust to explore curiosity-based scientific work rather than being strictly monitored administratively.

However, as widely agreed by the interviews, both types of tension are considered as potential concerns in MPG and HGF but as actual conflicts in CAS. The interviewed MPG scientists claim to be equipped with the best working conditions; they enjoy scientific freedom to explore curiosity-based science that is neither distracted by competitive research grants nor strictly monitored. Such a high level of scientific freedom at L3 at MPG matches its strict operational rules from L2 to L1, as discussed in section 4.3.3. Conversely, in HGF, strictly defined research fields and themes at L3 match the high level of operational autonomy at L2. The worst situation is that of CAS scientists because of their low proportion of block funding. They are found to be distracted and overburdened by both onerous control and administrative procedures when applying for the large number of complex and competitive research grants, and have very little scientific freedom.

The financial burden of CAS scientists is considered as the core reason of the above two types of tensions between L2 and L3. This argument is verified in the case study based on two interviews of the CAS president Bai conducted by Poo and Wang (2014) and by *Nature* (Cyranoski, 2014). As these two interviews present, because of the shortage of institutional block funding, all management overhead in CAS institutes—logistic service, building maintenance, electricity and water—is charged to research grants, which requires researchers to chip in their own resources (Poo a& Wang, 2014, p. 10). Furthermore, the overall proportion of industrial sponsorship in all the research grants is quite small (interview findings). Some underperformance results from low salaries for scientists in CAS institutes who scramble to supplement their incomes with multiple grants, which gives rise to their research being very fragmented (Cyranoski, 2014, p. 468). The interviewed

senior researchers in CAS are overwhelmed by the pressure of fighting for third-party research grants; as one senior researcher commented:

I always wish to invest more time and energy in laboratories rather than being busy travelling and in meetings for applying competitive research projects. I hope that one day the long-term stable funding will meet survival demands for researchers so that they will be relieved from busy grant applications and able to concentrate on their scientific work with much more attention than right now. (A CAS senior researcher)

Based on such financial burdens, there are related tensions identified concerning CAS scientists. The first one is their lack of *scientific collaboration*, as discussed in the previous section from the perspective of L2. At L3, according to the interview with the CAS president by *Nature*, ‘scientists shy away from collaborations because co-authorship dilutes their achievements in the eyes of grant committees’ (Cyranoski, 2014, p. 468). This causes much duplication of work and missed opportunities to share knowledge among CAS scientists at L3 and their embedded institutes at L2. A second relevant tension is the difficulty for early-career researchers and PhD students to produce indigenous innovation and scientific networking given their limited funding.

Another relevant tension in CAS is identified concerning the investment proportion between personnel and non-labour expenses in research institutes and programmes. According to the 2013 annual reports of the three PRIs, the proportion of personnel cost in the total expenditure of institution is 27.8% in CAS, lower than that in MPG (48%); there is no similar statistics in the HGF annual report. Many interviewed researchers in CAS were worried about the small proportion of investment in management and training of specialised technicians on some expensive laboratory equipment, which leads to low-quality operation and inefficient maintenance of the equipment. The following evidence is from explicit observations of interviewed researchers in CAS.

We have no special technicians to maintain scientific facilities; usually any research group who use them takes care. Yet, some contemporary users as well as the related training seem not very professional. (A CAS senior researcher)

Investment on expensive equipment is very much preferred over recruiting experienced technicians or training excellent ones to operate the equipment. Some equipment even stays unpacked since purchase because we do not have the expertise to take good care of it. This is really a huge waste. We know in Western countries, high-level and well-paid technicians are particularly responsible for operating and maintaining facilities to keep them continually working and extend their useful life as long as possible. We expect special institutional funding for training such technicians. (A CAS senior researcher)

Moreover, in CAS, the proportion of labour expenses in research programmes and projects is usually restricted below 30% and sometimes below 15%. Even with the increased institutional block funding, only about 25% was put into personnel, but over 70% into importing expensive equipment, which seemed more legitimate for the headquarters and the public funding bodies (interview findings). A potential reason could be assumed as the overall low level of scientists' income in China. Furthermore, the overall efficiency of using the expensive equipment is considered by the interviewees as very low because of the lack of professional technicians. Such investment preference for facilities rather than for personnel is considered by this study very negative for CAS research in the long term.

5.5 Governance Tensions on Strategy, Funding, and Operation

This section summarises the above specific governance tensions found at MPG, HGF, and CAS in terms of strategy, funding, and operation, as shown in Table 5.1 at the beginning of this chapter.

Strategy

Based on the empirical findings in sections 5.1 to 5.4, there are two main tensions related to strategy in the three PRIs. These two strategic challenges are discussed only between stakeholders at the highest levels of the organisational hierarchy (mostly at L1 and partly at L2). The first is between external justification of public funding and internal governance of heterogeneous research at L1. While the former asks for public understanding, support, and transparency of organisational governance, the latter is always complex, to some degree confidential and difficult for external stakeholders to understand. The conflicting expectations between

external and internal stakeholders within MPG, HGF, and CAS were presented in section 5.1. The complexity of internal governance of heterogeneous research is reflected by several tensions from L1 to L2 (section 5.2), among units at L2 (section 5.3), and from L2 to L3 (section 5.4). Therefore, the organisational strategies of PRIs need a balanced consideration of both external justifications by completing publicly defined missions and internal governance of heterogeneous R&D activities within different research disciplines, fields, and institutes.

The second tension about strategy deals with the conflicting goals between various units at L2 (institutes, centres, programmes, projects, and facilities) when they interpret overarching organisational strategies, always abstract, into concrete R&D tasks. As explained in sections 5.1 and 5.2, this tension takes place mostly in HGF and CAS, to different degrees, because their large numbers of legally independent units at L2 are required to have strong strategic alignment with the overall institutions. As explained in Chapter 4, unlike the legally dependent MPG institutes, all the HGF centres and many of the CAS institutes at L2 are legally independent with high levels of autonomous development. Even more difficult is the case of CAS where the complex and diverse research orientations have caused low efficiency of diffusion of strategies from L1 to L2. This led to a series of rigorous reforms, such as KIP introduced in section 4.2, impacting the research landscape of China.

Funding

The strong requirement for public justification results from the high level of public funding of the three PRIs analysed. As shown by Table 5.1, the common financial controversies at L1 in the three cases are about deciding internal resource allocation for the main funded subjects at L2—research institutes in MPG and CAS and large programmes across centres in HGF. As shown in section 4.2, these main funded subjects at L2 are both the key R&D producers and the evaluated subjects in the three PRIs. The principles of resource allocation are impacted by many factors; two of the commonly found factors in our cases are the scientific performance and potentiality of the funded subjects at L2. However, the challenge of standardising scientific performance and potentiality in heterogeneous research fields is confirmed by all interviewees. Therefore, financial tensions at L2 exist in all the three PRIs to different degrees in terms of funding priorities between institutes and programmes focusing on different research fields, also between ongoing

excellent research and strategic potential research that have not performed satisfactorily.

Across the three PRIs studied, there was little financial competition observed between fields or institutes of MPG, as shown in section 5.2. For both HGF and CAS, the conflict between collaboration and competition between research centres or institutes at L2 is quite obvious. The examples are presented in section 5.3. In addition, another conflict identified exists inside the units at L2 between their proportions of institutional block funding and competitive grants. This tension is more obvious in CAS than MPG and HGF because the high proportions of block funding in MPG institutes and HGF programmes have been very stable for decades. In contrast, the low proportion of block funding in CAS is considered as the fundamental cause of the mixed and fragmented research orientations (fundamental, strategic, applied) among the institutes.

However, third-party funding, with particular requirements for themes and applications, are found in the three cases to facilitate institutes' R&D diversity and autonomous development. This may entail the risk that PRIs' overarching organisational strategies not only become decentralised but also get spoiled. Because of that, third-party funding is only supplementary in a low proportion in the German cases, rather than the major resource. Yet, the relatively high level of third-party funding—over 50% and in some institutes nearly 70%—give rise to a financial burden of CAS scientists at L3 when applying for competitive research grants, with examples in section 5.4.

Operation

As Table 5.1 shows, the operational tension at L1 refers to the overarching policies that are not one-size-fits-all units at L2. This tension shares the fundamental reason with the other tensions at L1. That is, it is difficult for the top-down decision-making to be fair and effective for the large number of heterogeneous units at L2. These units are the scientifically autonomous MPG institutes, all the HGF centres, and some of the CAS institutes that are legally independent with operational autonomy.

As introduced in section 4.3.3, MPIs develop quite independently in scientific ways but they are legally dependent on operational rules made by the headquarters, such as the restriction of the number and budget for regular staff positions. In

contrast, although conducting research in strictly defined fields, the legally independent HGF centres are operated more independently. For example, their external collaboration is more guided by the strategies of the centres than by the whole institution. Some of the 104 CAS institutes have legal status while some do not. Many practical reasons (history, geography, number of personnel, etc.) cause such a gap between the institutes in terms of quota restrictions on applications for internal resources of CAS. Its ongoing reforms, such as PAP, aim to bridge the gap.

Furthermore, the operational tension at L3—scientific freedom threatened by scientists' burden and distraction when applying for competitive funding—is widely discussed in the interviews. This tension is considered as a potential concern in MPG and HGF, but as an actual tension in CAS. Some other related problems for CAS scientists, such as the lack of collaboration, are discussed in section 5.4.

To sum up this section, Figure 5.1 visualises some governance tensions with two key conflicting situations at MPG, HGF, and CAS. Within each tension between the conflicting situations, the positions of the three PRIs are interpreted by this thesis in a comparative way based on the individual case studies. Only the tension between institutional block funding and competitive project funding has explicit quantitative evidence. All the other conflicts are interpreted from qualitative evidence from the interviews. In fact, this figure selects the most significant tensions identified with only two key conflicting situations. There is one important tension discussed in this chapter but excluded in this figure—the overarching policies at L1 are not one-size-fits-all units at L2. This tension has multiple conflicting situations with very complex reasons and could not be measured comparatively among the three PRIs.

Moreover, some gradual changes over time have been found, as the arrows show, from one situation to the other. Most of the changes between the conflicting situations were observed in CAS due to its vigorous reforms. MPG and HGF are performing more stably than CAS in the conflicting situations. Their governance has not changed significantly since their establishment. The changes in operational aspects without conflicting situations, such as increasing external collaborations, are not presented in Figure 5.1. For CAS, its governance dynamics along with the large-scale and costly reforms are easier to trace. For instance, CAS's proportion of institutional block funding has been very slowly increased over the last decade.

The Balancing Role of Evaluation Mechanisms

Figure 5.1 Key governance tensions with two conflicting situations

Source: the author, 2014

To conclude this chapter, specific governance conflicts within MPG, HGF, and CAS in terms of strategy, funding and operation and taking place across the three organisational levels (L1, L2, L3) are identified. Some tensions are shared by the three PRIs with different positions between the conflicting situations, as shown in Figure 5.1; some tensions are unique. For the shared conflicts, the most significant reason found is the heterogeneity of the key research units at L2—institutes in MPG and CAS, and centres and research programmes in HGF. The unique tensions are in line with the particular characteristics of organisational governance as discussed in Chapter 4, such as the unbalanced funding proportion between scientists' income and expensive equipment in CAS. The empirical evidence validated the enriched framework proposed in the previous chapter. The next question is how to balance the governance tensions, more precisely, how the evaluation mechanisms of MPG, HGF, and CAS respond to their particular governance tensions. This will be analysed empirically in the next chapter.

6 Evaluation Mechanisms Respond to and Facilitate Organisational Governance at MPG, HGF, and CAS

How do the evaluation mechanisms of Max Planck Society (MPG), Helmholtz Association (HGF), and the Chinese Academy of Sciences (CAS) respond to governance tensions and facilitate the organisational governance? This chapter contributes to answering this overarching research question by looking at the linkages between organisational governance and evaluation mechanisms proposed in the basic and enriched conceptual frameworks (Figures 2.1 and 4.3). Section 6.1 introduces the specific key evaluation activities composing the overarching evaluation mechanisms of the three publicly funded research institutions (PRIs). Comparative analyses of the three evaluation mechanisms are presented in section 6.2 in terms of integrated indicators reflecting governance requirements. Section 6.3 shows the balancing processes of the evaluation mechanisms in two functioning way—institutionalisation (section 6.3.1) and stakeholders' participation and interactions (section 6.3.2) reflecting the governance conflicts proposed in Chapter 5. Section 6.4 discusses how the evaluation mechanisms facilitate organisational learning by providing evaluation findings as knowledge bases and promoting stakeholders' negotiations in evaluation processes. The balanced governance results that stem from the functions of the evaluation mechanisms are discussed in the last section. All these findings at MPG, HGF, and CAS contribute to the general understanding of the implication of evaluation mechanisms as governance instruments, as will be discussed further in Chapter 7.

6.1 Evaluation Mechanisms at MPG, HGF, and CAS

This section describes the characteristics of the key evaluation activities that together compose the evaluation mechanisms of MPG, HGF, and CAS. As defined, *evaluation mechanisms* of PRIs include all types of individual evaluation activities and operate vis-à-vis the complex institutional environments and stakeholders. Yet, this section focuses on the key evaluation activities that are more closely related to the organisational governance of the three PRIs as discussed in the previous chapters. These activities are described in annual reports, reports on evaluation procedures of the three PRIs, and during the interviews. More supplementary details of their evaluation practices can be found in the sections of evaluation practices in the appendices.

6.1.1 Evaluation Mechanism at MPG

Figure 6.1 gives an overview of the four main types of evaluation activities at MPG as well as their coordination over time. Overall, this mechanism is found to be able to cope with a diverse spectrum of research activities to guarantee their leading fundamental research. MPG claims that the best research and scientists can only be evaluated by the best peers worldwide. To be accepted as critical observers, the evaluators recruited are experts with high reputations and rich experiences within particular fields whose own research puts them on par with the Max Planck Institute (MPI) directors. The four key types of evaluation activities composing the mechanisms are briefly introduced one by one.

Figure 6.1 An overview of the MPG's evaluation mechanism

Source: the author, 2013

Ex-ante evaluations of the scientific achievements and the personality of candidates for appointments of scientific leaders are greatly valued by each PRI studied. However, only MPG prioritises this evaluation as the most important type within the overarching mechanism. The *ex-ante* evaluation in MPG is considered by this study perfectly in line with its person-centred governance because it is the one and only evaluation activity that directly decides on the funding of MPIs. The establishment of new institutes and the directorial appointments are built on highly strict and confidential evaluations, by international experts, of both individual scientists and the entire scientific concept of the institute. The decisions

are made only when the leaders with the highest competence, originality, and creative potential in their fields and the future development of the institute are convincing for all the scientific members (see section 4.4). According to interviews, once appointed, the new directors, as well as scientific members, enjoy continual block funding of the institute and a high level of freedom to determine detailed research subjects, structures, and personnel within ideal working conditions. The *ex-ante* evaluation for appointing new scientific leaders is considered by the interviewees from the MPG headquarters as the most powerful tool to guarantee the effects of Harnack Principle with high levels of block funding and scientific freedom for MPG scientists.

Ex-post institute evaluations aim at research improvement in individual MPIs every two years. The evaluators are members of the Scientific Advisory Board (SAB) of each MPI, usually 8–10 international experts, 99% outside MPG and 85% outside Germany, after strict selection and rotation procedures, according to the Rules of Scientific Advisory Boards (2015), hereinafter referred to as SAB rules (Appendix A4). The SAB members of each MPI professionally evaluate and advise the existing research areas, performance and potential to ensure they are addressed and developing at the highest international level. Such frequent and professional evaluation is considered to guarantee the quality of MPG research as well as the freedom and trust of MPG scientists. Recommendations are also about more efficient management and resource allocation among the internal groups at each MPI. As discussed in section 5.2, in principle, any institutes that no longer meet the high expectations of SAB are either restricted or closed. However, no MPI has ever been closed because of a negative evaluation. As one interviewee at the headquarters posited, frequently evaluating MPIs at a high and international level helps to prevent such a ‘surprise’ happening. The closure consequence is more likely to happen at the level of a single department.

Extended evaluations, based on individual institute evaluations, are conducted every six years by the Research Field Committee on the scientific clusters of the 82 MPIs. This evaluation aims to assess the institutes’ performance and plans, from a medium-term perspective, to promote collaboration and future synergies between MPIs in close research fields but that are developing autonomously. Usually two or three external rapporteurs visit the SAB meetings of all the institutes clustered in one field and give feedback on their evaluation procedures and results (SAB rules, 2015). The two rapporteurs chosen are not only leading scientists but also have a

particularly deep understanding of the German research and development (R&D) context. This is seen as an effort by MPG to respond to national R&D needs.

Strategic bibliometrics has an *ad hoc* feature and is conducted on a voluntary basis for the various evaluated subjects (research topics, fields, individual researchers, institutions, countries, etc.). This evaluation interprets quantitative bibliometric data from publication databases and proposes qualitative and comparative analyses of the evaluated subjects at many levels. The standardised approaches of this evaluation have been published in academic journals, and this evaluation supplements other evaluations by providing comparative and strategic application. For example, bibliometrics experts analyse thoroughly the quantity and quality of published journal papers of one MPI, and consider particular contexts (fields, disciplines, scientists, team structures, facilities, development scales and speed, etc.), to check whether this MPI belongs to the top one percent of superior institutes within the scientific community. Such information is considered by the headquarters to be of high standard, valid, trustworthy, and confidential. It is used for policy-making at both the institute and MPG levels by the high-level leaders to attract leading scientists. Moreover, comparative evidence yielded by this evaluation is found to contribute to both the evaluation of individual MPIs,²⁴ and the extended evaluation about research fields.

The above four types of evaluation activities are integrated into a MPG evaluation mechanism which will be compared with HGF's and CAS's mechanisms in section 6.2. The two key evaluation activities within HGF will be introduced in the next subsection.

6.1.2 Evaluation Mechanism at HGF

HGF has two main types of evaluation activities—strategic **Programme oriented Funding (PoF) evaluation** at programme level, which includes the evaluation of the large-scale infrastructures of HGF centres, and in-depth scientific **midterm review** at centre level. As discussed in section 4.2, PoF requests all the centres to apply for their funding (almost 2/3 of the total amount) for the next five years by coordinating programme proposals to be evaluated by international experts. The descriptions of specific procedures, criteria, and seven-score rating system for

²⁴ Strategic bibliometrics is only applicable for natural sciences, not the social sciences or humanities.

programmes and the subordinate topics can be found in Appendix B4. At the midterm of this five-year period, each centre organises an internal review concentrating on its own R&D and development. These two types of evaluation working at different levels complement each other chronologically, as shown in Figure 6.2, and substantially in the following aspects.

In terms of purposes, PoF evaluations decide on the strategic planning, allocation and control of the core budget of the six research fields and the involved programmes and centres for the next five years. Strategic goals in each field are prioritised over R&D performance of the programmes and centres in PoF evaluations to make sure the potential of the programmes translate research into beneficial solutions to societal challenges. Yet, the midterm review at each HGF centre assures the R&D quality and development at the centre level.

Figure 6.2 An overview of HGF's evaluation mechanism

Source: HGF headquarters, 2013

Both types of evaluation depend on international review panels with different evaluators, budgets, and highlights. In each PoF round, there are around 15 reviewers for each programme, and evaluation of 30 programmes cost about €2 million in total (interview findings). Two-day on-site PoF evaluation of individual programmes usually takes place at the predominant coordinating centre. Which

centre is selected as the PoF evaluation location leads to discussions among the participating centres because ‘people always want to have this on-site review at their own workplace’ (interview findings). One key reason could be that the stakeholders’ negotiations on PoF evaluations mostly take place at the selected centre. Each programme proposal is evaluated looking at strategic goals and challenges, justification of public resources, contributions to the field, the quality and productivity of R&D units, and training activities. Each research topic is considered as the smallest evaluation unit without discussing performance of individual scientists.

Midterm reviews are funded and organised by each centre itself without standardised procedures. As a case study of the large-scale HGF centre X shows, with seven programmes across fields, its midterm review is usually more complex than separated programme reviews, and gives a more complete picture of the centre for internal resource allocation. Midterm reviews usually take more time—often a month—than PoF evaluations, and invite more evaluators, over 100, some of whom are also members of PoF evaluation teams. Institutes within one centre and other internal groups divided by research topics are reviewed scientifically. The related PoF recommendations are also referred to. In this way, centres have power and space to discuss their own strategies. The headquarters are informed of the midterm results which are also shared with the next PoF proposals.

As for impact, midterm reviews produce both scientific and governance recommendations for individual centres. Research excellence justified by midterm reviews helps the centres to apply for more funding in the next PoF round. Usually, results of the PoF evaluation and midterm review go towards the same direction—as one of the interviewees in HGF centre X put it:

Decisions to change significantly the field or orientation in one PoF round are unlikely to happen. This is quite nice and easy for us. Based on our midterm evaluation, the PoF evaluations usually say you are good and please continue. (A scientist of HGF centre X)

The PoF evaluations have direct impact on the amount of funding for each programme and its ratio of resources from the federal and state governments. As agreed within each programme proposal, the participating centres interpret the PoF evaluation results and use them to negotiate their specific budget for the next five year. Reviewing the past three PoF rounds, overall, programmes are

consistently receiving increased funding of 2%–5% for each five-year period. However, within each programme, the allocation among participating centres and topics are adjusted according to detailed recommendations from the PoF evaluation.

Responding to the financial tension between units at L2 as discussed in section 5.4, PoF evaluations are responsible for providing valid and trustworthy evidence to answer the conflicting questions, as explained by a top executive at HGF headquarters:

Financial increase by 2%–5% in one programme could mean about €13–€14 million additional funding for one centre. It is a lot of money and the allocation is very meaningful if one institute or topic gets 10% more but others get 5% less. Not all the institutes and their research topics can win at the same time but they all trust PoF evaluation results to guide resource allocation. (Executive at HGF headquarters)

Each interviewed centre claims to integrate the scientific and governance recommendations in terms of strategy, funding, and operation, from both the PoF evaluation and midterm reviews for its own development. For example, the aspect of operation of individual programmes evaluated refers to research collaboration, talent training and infrastructures, which do not carry as significant weight for the final evaluation results than the strategy aspect. The effect of integrating the two main activities as an overarching evaluation mechanism will be analysed in the following sections.

6.1.3 Evaluation Mechanism at CAS

As the case study shows, evaluation at CAS started late and developed very slowly and was fragmented for a long time. The historical development of R&D evaluation in CAS and some specific evaluation activities can be found in Appendix C4. Overall, evaluation activities at CAS about research programmes, projects, talent, and laboratories have been independent each other with very limited strategic impact. These different evaluations are undertaken by different units—sometimes different departments of the headquarters—that do not share the evaluation criteria and results. Furthermore, the different evaluation organisers have little awareness of the need to coordinate and integrate these evaluations. The independence and lack of coordination of the chaotic evaluation activities at CAS resulted in its lack of an

overarching evaluation mechanism for a long time. However, in 2012, and realising such problems, CAS decided to create a new evaluation mechanism to aggregate the fragmented evaluation activities—**One–Three–Five evaluation**²⁵ as shown in Figure 6.3.

Figure 6.3 An overview of CAS's evaluation mechanism

Source: the author, 2014

Given CAS's large number of short-term competitive project funding of CAS as discussed in section 4.3.2, the problem of fragmented project evaluation, both internally and externally, is considered by interviewees to be significantly negative for the development of CAS research. The peer reviews of these projects prefer to look at quantitative indicators and applicants' networks rather than the long-term potential of research. This has provoked inappropriate behaviour like aggressive impetuous attitudes, even misconduct like plagiarism and data manipulation. Moreover, the problem of aggressive evaluation criteria was acknowledged generally by the CAS's president (Bai, 2007, p. 109)—'emphasising quantitative indicators but ignoring qualitative contribution and capacities, and the undervaluation of societal impact'. As Michelle (2015, p. 13) states, an academically fair, objective, and transparent evaluation system is urgently required for CAS to give young and promising researchers stable support and confidence, and

²⁵ By mid-2015, the One-Three-Five evaluation was finished in all the 104 institutes with a comprehensive performance audit, based on their annual key performance indicators proposed in the mission statement document of each institute. However, no report or article has yet been published about the effect.

therefore to guarantee the excellence of research and attract outstanding potential candidates.

Institute evaluations are found to be the most influential and strategic evaluation forms that promote an objective and transparent evaluation system at CAS. Institute evaluations have been institutionalised in a top-down way by a series of reforms, which is elaborated in section 6.4.1. There was no standardised evaluation of all the CAS institutes before 2012 when the *One–Three–Five institute evaluation* started. Until now, institutes' annual reports, which are prepared at L2 and commented at L1, have been considered as the most normalised approach of supervision from L1 to L2. Annual reports summarise R&D performance, management, and administrations, usually in a descriptive way. Based on that, some overarching policies are made by the headquarters. However, some pioneering institutes have constantly and autonomously undertaken international evaluation of their own development, such as the CAS institute of Neuroscience in Shanghai. These *ad hoc* international evaluations are organised by individual scientists with excellent reputations and international networking. These evaluations are considered by the evaluated institutes as opportunities to learn international experiences and recommendations that will impact the institutes' strategic plans and internal resource allocation.

Since 2012, the newly developing **One–Three–Five evaluation** aimed at standardised evaluation for all the institutes is considered by this thesis as the CAS's evaluation mechanism. The evaluation was initiated as a response to the Major Outcome oriented governance of CAS institutes to fulfil *Innovation 2020*. The Major Outcome principle guides the ambitious target of *One–Three–Five evaluation* to aggregate the existing various evaluations in individual institutes (Bai, 2012). According to the report from CAS Evaluation Centre²⁶ (2013), the Major Outcome of each institute is promoted by assessing its one position, three breakthroughs and five key fostering directions every five years. As presented in section 4.2, *one position* of an institute means its current strengths and ranking in upper (or lower) echelons in the scientific communities, including its long-term target. *Three breakthroughs* refer to the most promising R&D targets within the institute to achieve its future position. *Five key fostering directions* are specific research topics where most of the existing resources are invested to realise the breakthroughs.

²⁶ Evaluation Research Centre of CAS is described in Appendix C4.

Instead of overarching evaluation from L1 to L2, the innovative *One–Three–Five evaluation* targets aggregating the fragmented research and resources of the CAS institutes and promoting their *Major Outcomes* (Bai, 2012). Each institute signs a mission statement including specific targets that are confirmed by the headquarters and assessed strictly by international experts every five years. Reviews and suggestions from evaluators concentrate on the focused *Major Outcomes* rather than on the comprehensive performance of institutes. There are relevant policies announced to support this evaluation, such as simplified administrations, more empowerment of institutes on interpretation of their One–Three–Five aspects, and strategic planning. Therefore, institutes are more empowered to arrange their resources flexibly for other tasks beyond their One–Three–Five–related missions.

The responses from the interviewees about this type of evaluation can be found in Appendix C4, such as the following two short examples:

It is much better that the institutes interpret themselves the meaning of Major Outcome rather than the headquarters. Each institute interprets the One–Three–Five evaluation policy and responds according to its own characteristics. A big centre may extend to One–Six–Ten aspects and a small one only makes two breakthroughs and four directions. (Director of CAS Institute a)

The headquarters could not possibly view all the research work of all institutes, so it focuses on the expected Major Outcome. The One–Three–Five evaluation is to find out and avoid the homogeneous competition among institutes and to promote significant outcomes of the whole CAS. The increased block funding is promised for the expected breakthroughs and directions if you pass the evaluation in five years. (Deputy of R&D Department of CAS institute c)

An additional innovation of the One–Three–Five evaluation is the increasing consideration of various research orientations and the related involvement of international experts. Guidelines on expert selection highlight different criteria for considering research orientations. For instance, experts in fundamental research are mostly international peers with outstanding reputations. Experts in strategic or applied research include more domestic specialists and users with rich local experiences.

The above individual evaluation mechanisms at MPG, HGF, and CAS will be integrated into analyses of their role responding to governance requirements as discussed in the next section.

6.2 The Evaluation Mechanisms Responding to Governance Requirements

Based on the individual case studies, Table 6.1 offers an overview of the characteristics of various evaluation activities at MPG, HGF, and CAS through the nine selected dimensions proposed in section 3.2. This table provides the basis to understand the *systemic nature of evaluation mechanisms* composed by those various evaluation activities, as conceptually discussed in section 2.3.1. In practice, individual evaluation activities in each studied PRI supplement each other by contributing wide-ranging characteristics on the dimensions to make the overarching mechanisms comprehensive. Therefore, **the overarching evaluation mechanisms can help to respond to the comprehensive governance requirements.** For example, different evaluation activities taking place at different points in time could be arranged systemically by the evaluation mechanisms to respond to various financial decisions; *ex-ante* evaluations legitimise funding plans towards MPG institutes permanently and HGF programmes in the next five years; *Ex-post* evaluations justify financial adjustment of institutional funding and resource allocation; other regular and *ad hoc* evaluations provide evidence for funding arrangements in relation to talent, infrastructures, and third-party projects. More examples on this core claim will be presented throughout this chapter and based on the characteristics of individual evaluation activities summarised in Table 6.1.

Overall, almost all the dimensions and characteristics in Table 6.1 are shared by the evaluation mechanisms of the three PRIs, including both the ticked characteristics and the explanations about specific activities. The shared dimensions and characteristics of evaluation activities, particularly the wide-ranging levels, subjects, timing, and participation, are common for PRIs beyond the three cases. The different characteristics, the blank boxes in Table 6.1, will be discussed with examples below in this section. But, this section does not discuss the items in this table one-by-one repeating some analyses done in section 6.1 on the dimensions of *level*, *subject*, and *timing*. Some supplementary information can be found in the appendices. The dimensions of *participation* and *use* will be analysed in relation to the governance in the following sections.

The Balancing Role of Evaluation Mechanisms

Dimensions	Characteristics		Various evaluation activities composing the evaluation mechanisms		
			MPG	HGF	CAS
Level	System	National	√		
		Institutional	√		
	Institutes		√		
	Programmes and projects		√		
	Individuals		√		√
Subject	Internal units		Institutes; clusters	Centres; programmes	Institutes; programmes
	Irregular external projects		√		
Timing	Ex-ante		Appointment of leaders	PoF evaluation	One–Three–Five preparation
	Ex-post		Institute evaluations		
	Ad hoc		Peer review of competitive project funding		
	Regular		Evaluation of personnel, infrastructures, etc.		
Participation	Internal scientists		√		
	External experts		Many: politicians, academics, businessmen	A few	
Use	Research quality assurance		√		
	Strategic and financial decisions			√	√

Table 6.1 Characteristics of the evaluation mechanisms of the three PRIs

Source: the author, 2014, based on empirical evidence; the table structure comes from ‘The Practice of Evaluation in Innovation Policy in Europe’, Edler et al., 2012.

Based on the similar systemic nature—composed of various characteristics of individual evaluation activities—the **three overarching evaluation mechanisms respond to their PRI’s own governance requirements differently** as shown in the spider chart (Figure 6.4). Seven indicators in this figure illustrate the different characteristics of the three evaluation mechanisms that are related to the governance characteristics discussed in Chapter 4 and to some governance tensions analysed in Chapter 5. The seven indicators in the spider chart focus on the impacts of evaluation mechanisms received and produced in relation to governance. For example, the levels of *strategic use* and *funding use* of the evaluation mechanisms explicitly reflect how they respond to governance requirements in relation to strategy and funding. Specifically, both high-level performances on these two indicators of the HGF evaluation mechanism result

from its programme-centred governance—only the research programmes strictly follow the strategic guidelines, and they can be funded consistently. Overall, the qualitative characteristics of the three evaluation mechanisms, reflected by the seven indicators, were measured and compared in a quantitative way in Figure 6.4, based on the policy documents and interviews at the three PRIs.²⁷ The outer rings refer to higher scores of the indicators in the spider chart; the inner, the lower. Yet, such comparison does not indicate any specific scores but presents their general differences on the indicators.

Figure 6.4 Characteristics of the three evaluation mechanisms responding to the governance requirements

Source: the author, 2014²⁸

Why do the evaluation mechanisms of MPG, HGF, and CAS perform three distinct profiles as shown in Figure 6.4? The reason is based on a similar principle—the evaluation mechanisms actively respond to the *different* organisational missions

²⁷ Ad hoc evaluations (projects, personnel, infrastructures) are not included as they are not comparable. The selection of the indicators comes from 'The Practice of Evaluation in Innovation Policy in Europe', Edler et al., 2012.

²⁸ Figure 6.4 was presented and discussed at several international conferences, such as the European Evaluation Society Conference 2014 in Dublin, Ireland.

and governance requirements of each institution. The following explanations are briefly presented case by case; more examples will be shown in next section on the balancing processes of evaluation mechanisms.

MPG

The MPG evaluation mechanism focuses on justifying its mission of *doing the best fundamental research* and ensuring a high level of scientific freedom and trust. On the one hand, to ensure research excellence, the research evaluations are carefully organised in terms of frequency and large-scale issues (selection and rotation of international experts), as introduced in section 6.1. For example, the selection of institute directors is based on strict *ex-ante* evaluation and their work is *ex-post* reviewed by SAB members every two years and compared with other institutes in the extended evaluation every six years. On the other hand, a high level of trust of MPG scientists is guaranteed because the impact of the research evaluations concentrates on research itself rather than governance aspects. For example, the evaluated institutes and scientists receive suggestions centred on research improvement, which has much less impact on the strategy, funding, and operation of the institutes at L2 and the institution at L1 at MPG than at HGF and CAS. Furthermore, there are relatively fewer evaluation guidelines from stakeholders on the governance aspect of strategy and funding at MPG than at HGF and CAS. Yet, the core evaluation guidelines—SAB rules²⁹—are published on the MPG website for the purpose of public justification, which is the most transparent within the three cases. However, evaluation results can be accessed by the fewest stakeholders, which is seen as *the most confidential* feature in Figure 6.4. Different versions of evaluation results at MPG will be discussed in section 6.3.2 related to stakeholders' interaction. Overall, the three PRIs all agree, according to interviewees, that the reason for keeping evaluation results confidential is their uses for internal governance.

HGF

As shown in Figure 6.4, the overall shape of the HGF evaluation mechanism performs quite oppositely to that of MPG on several indicators. Both of the evaluation activities at HGF—PoF evaluation across all centres and midterm review

²⁹ The SAB rules, as introduced in section 6.1, include selection and rotation procedures of SAB members and general evaluation criteria for research quality.

at each centre—are less frequent than at MPG, but more strategic and financial impacts stem from them in relation to governance. Research programmes of HGF centres have been oriented by thematic and strategic goals indicated by two key evaluation guidelines—Strategic Guidelines and Position Papers.³⁰ These guidelines involving multiple stakeholders specify detailed scientific objectives, budgetary frameworks, and evaluation criteria and are strictly implemented in a top-down way, which leads to a lower level of scientific freedom and trust. The contrasting situations of MPG and HGF support the claim that the different evaluation mechanisms comply with their respective organisational missions and the related governance requirements.

CAS

In general, CAS exposes a rather small-scale profile on most of the indicators in Figure 6.4, as the red shape shows, mainly because its key evaluation activity—*One–Three–Five evaluation* is still developing. The other, earlier evaluation efforts have shown only little governance relevance, as introduced in the previous section. Many of the indicators in Figure 6.4 are increasingly impacted by the *One–Three–Five evaluation*, such as developmental autonomy at L2 and scientific freedom and trust at L3. The institutionalisation of the evaluation mechanism at CAS will be analysed in section 6.3.1. One can assume that over the next 15–20 years, the red shape will probably expand quickly with much higher performance on the indicators than now.

Overall, the CAS governance target of aggregating Major Outcomes of institutes relies on the definition and evaluation of the aggregated *One–Three–Five aspects* of each evaluated institute. Strategic planning and financial allocation based on evaluation evidence at both L2 and L3 are increasingly oriented by the institutes' long-term *one position, three breakthroughs and five key fostering directions* rather than by the previous fragmented short-term research projects. Larger-scale and more international evaluations are taking place to increase the institutional block funding, which helps to relieve the financial burdens of CAS scientists as proposed in section 5.5. The openness and transparency of the evaluation guidelines is expected, by the author, to be strengthened by the increasing application of international norms and standards in CAS's evaluations. The core governance

³⁰ The two guidelines were introduced in section 4.2 and the supplementary information can be found in Appendix B3.

targets—aggregated strategies and progressively increasing block funding—become the key concerns of the evaluation mechanism based on, but far beyond, the previous concentration on research quality.

This section presented the characteristics of the evaluation mechanisms of MPG, HGF, and CAS and their systemic natures and capabilities to respond to governance requirements. The findings confirmed the organisational use of evaluation, conceptually discussed in sections 2.3, and how governance impacts on the design of the evaluation mechanisms on particular indicators. This section thus provides a basis for interpreting the processes of the evaluation mechanisms as reflecting and balancing the governance tensions identified in Chapter 5. This will be discussed in the next section.

6.3 The Balancing Processes of the Evaluation Mechanisms

Since the evaluation mechanisms have been designed according to governance requirements at MPG, HGF, and CAS as discussed in section 6.2, a resulting question is how do they become instruments to balance governance tensions identified in Chapter 5, concerning the three governance aspects and across the three organisational levels? According to the interviews at the three PRIs with people who specifically focused on mitigating governance conflicts, both actual and potential tensions are continuously being analysed for effective solutions. Their experiences of overcoming the tensions could be summarised as two approaches: a) institutional reforms on governance arrangements, such as development of HGF programmes and change of research structure in CAS institutes, and b) learning about governance from international experiences at successful PRIs. In both approaches, evaluation is considered as a core concept and a reliable and powerful instrument for governance as it provides evidence for relevant decision-making. This section explains the two ways of the balancing processes—the institutionalisation of evaluation mechanisms, to be elaborated in section 6.3.1, and evaluation mechanisms as facilitating participation and interaction vis-à-vis stakeholders around their diverging interests, to be explained in section 6.3.2.

6.3.1 Institutionalisation of Evaluation Mechanisms at MPG, HGF, and CAS

As introduced in section 2.2.2, the institutionalisation of evaluation mechanisms at PRIs in this thesis is defined as ***the process of professionalising and systemising various individual evaluation activities with consideration of and uses for organisational governance***. Reviewing the evaluation dynamics of the three PRIs in

the past 20 years, one could conclude that the three PRIs have professionalised and enriched their individual evaluation practices in terms of the dimensions presented in Table 6.1—levels, subjects, timing, participation, and use. The enriched dimensions of individual evaluation activities have consistently contributed to the overarching evaluation mechanisms responding to the governance requirements, as shown by the spider chart Figure 6.4. Such processes are interpreted by this thesis as the PRIs' ways to institutionalise the *evaluation mechanisms* centred on the systemic nature and the aggregated impact on the governance.

Figure 6.5 A general process of the institutionalisation of the evaluation mechanisms

Source: the author, 2014

Figure 6.5 shows three stages of a general institutionalisation process of the evaluation mechanisms since the 1990s in terms of the increasing concerns related to the organisational governance. These concerns, from basic to complex ones with longer-term and wider-range considerations, are reflected by the totality of the purposes, procedures and uses of individual evaluation activities. Overall, the evaluation mechanisms of the fast developing PRIs increasingly concern the governance requirements, specifically, the increasingly disciplinary fundamental research at MPG, the more professionally organised programmes at HGF, and the growing integration of R&D activities at CAS. Therefore, the overarching mechanisms have developed more comprehensively, e.g. in terms of level and subject, and more responsively towards governance requirements, e.g. through the use of evaluation results for strategic and funding decisions. More importantly, the

institutionalised evaluation mechanisms are more capable of balancing governance tensions by increasingly involving the conflicting issues, such as the autonomy of the evaluated institutes and their synergies at L2, and scientists' freedom and trust at L3. For example, evaluation activity I and activity II of the overarching evaluation mechanisms provide evidence respectively for the two conflicting situations of one tension as depicted in Figure 5.1 (section 5.5). That evidence could be systemically analysed to contribute solutions or mitigation of the tensions. More concrete example will be given within the three PRIs in this section below.

Looking at this general trend, the **distinct institutionalisation** of the evaluation mechanisms **at individual PRIs** are shown by in Figure 6.6. Indeed, with the same horizontal and vertical variables as Figure 6.5, Figure 6.6 does not compare the numbers of the dimensions in the vertical axis but presents their individual dynamics about the increasing speed (curves) and enlarging scale (bubbles) of the institutionalisation processes. Overall, MPG shows the flattest trend in the three cases, which means the least change was observed on both speed and scope of the evaluation institutionalisation. MPG started its research evaluation at a high level very early—in the 1970s—and keeps updating them without significant changes, which keeps them stable. In contrast, HGF established the core PoF evaluation quite late—in 2002—which is why its trend started from the x axis, but since then the evaluation has been institutionalised quickly with considerable effect. CAS started research evaluation from nearly zero in the 1990s, and since then has experienced various phases. However, since 2012, its new developing *One–Three–Five evaluation mechanism* has developed quickly. Specific explanations of the three cases will be presented in the following analyses.

Figure 6.6 Trends of institutionalisation of evaluation mechanisms in the three cases

Source: the author, 2014

How could the above interpretation of the trends of evaluation institutionalisation in the three cases and their comparisons be validated? To clarify the institutionalisation of evaluation mechanisms case-by-case, it is important to specify their typical elements found in the case studies, as shown in Table 6.2. To interpret the abstract concept of *institutionalisation of the evaluation mechanisms*, the elements in Table 6.2 found from interviews are often referred to by the interviewees when they review the past experiences as well as anticipate promoting the evaluation institutionalisation in the future. Furthermore, these practical elements are closely related to the elements of evaluation institutionalisation as conceptually discussed in section 2.3.2. In particular, the discussed *evaluation culture* at the institutional level is validated at the three PRIs as such culture is embraced by almost all of interviewees and the below elements.

Key elements during institutionalisation	The three PRIs studied		
	MPG	HGF	CAS
Quantity and quality of evaluation guidelines (such as policies, handbooks)	SAB rules become standardised and updated year by year	Strategic Guidelines and Position Papers become more comprehensive and specific round by round	Evaluation policy documents have more open targets and procedures
Involvement of international evaluators	Always high level	Intermediate	Steadily increasing from very few
Acceptability of evaluation results	Evaluations bring about critical comments and research recommendations	Evaluations bring about stable increase of programme funding	Evaluations bring about administrative procedures
Internal office responsible for organisation and coordination of evaluation activities	Institute Development and Evaluation	Research Section and research centres	Evaluation Centre

Table 6.2 Dynamics of the institutionalisation of the evaluation mechanisms

Source: the author, 2014

To look at the elements in Table 6.2 one by one, evaluation guidelines are the most accessible and transparent evidence to be referred to in the interviews. They usually specify procedures and instructions on how to contribute evidence for governance decisions. Most of those important guidelines at MPG, HGF, and CAS are updated after each round of evaluations. Next, involvement of international

experts is considered by all the three PRIs as an explicit contribution for the *evaluation institutionalisation* because it addresses global scientific concerns and keeps evaluations objective, especially when facing competition between the evaluated subjects. The acceptability of evaluation results within the three PRIs involves their different impacts towards and responses from the institutes at L2 and scientists at L3.

Next, each PRI studied has special internal unit that is responsible for organising and coordinating evaluation activities, sometimes even analysing evaluation results and impacts. They are all within the headquarters—Institute Development and Evaluation at MPG, Research Section at HGF, and Evaluation Centre at CAS—and claim to be independent. The institutionalisation processes of evaluation mechanisms at the three PRIs are mainly traced from the published policy documents and the interviews at these units. The dynamics of these units reflect the professionalisation of their respective organisations and analytic work related to the evaluation. For instance, the Evaluation Centre at CAS has expanded considerably in recent years because of the increasing evaluation activities taking place at CAS. The Institute Development and Evaluation at MPG and Research Section at HGF are more interactive with other departments within their headquarters to coordinate evaluation activities. Their interaction with other stakeholders is closely related to the internal diffusion of evaluations, which will be discussed in the next section. The following analyses will be about the individual institutionalisation processes of the three PRIs which started at different times and proceeded at varying speeds.

MPG

The reason why MPG has the *most stable institutionalisation trend*, in Figure 6.6, is that the governance considerations within the evaluation mechanism are limited. For its main mission of the *leading fundamental research*, the key evaluation activities at MPG focus on research improvement by attracting the world's leading scientists as evaluators and frequently have them update the evaluation criteria regarding knowledge advancements. The guideline updates for the core evaluation activities—*ex-ante* evaluation of institute directors and *ex-post* institute evaluation—are mostly in terms of recruitment requirements and evaluation methodologies, and with little strategic or financial uses for the evaluated units. The other two evaluation activities—extended evaluation for clusters and strategic

bibliometrics—have been developing with increasing strategic uses but are still much less institutionalised than the core evaluation activities.

The governance tensions identified at MPG are centred on the difficulty of ensuring that the complex research decisions at L2 and L3 are made appropriately, as discussed in section 5.5. To respond to that, the *evaluation culture* promoted within MPG is the **high level of acceptability and credibility of evaluation results** because the results come from the *best scientists* within particular research fields. As long as this strategy to attract the best scientists to evaluate research at MPG is followed, there will be no major changes of the stabilised evaluation mechanism. That is why, according to one interviewee, ‘the successful story of MPG evaluation is winning and promoting the best scientists all over the world’.

A related question at MPG is about the objectivity of these leading scientists as evaluators because usually there is very limited number of top experts within highly specialised research fields. The MPG interviewees agree that it is difficult to keep a distance between the evaluators and the institutes. However, it is not a problem but instead promotes the scientific trust of the evaluated research and the scientists, as commented by a high executive in MPG headquarters:

It is more worthy and convincing for us to have trustworthy colleagues from the small top communities as professional evaluators who probably know each other rather than some unknown experts unfamiliar with the fields.

To sum up, the highly stable evaluation mechanism at MPG results from its principle to recruit the best scientists as evaluators, which follows closely its mission and strategy of organisational governance, as presented in section 4.2. This also reflects that the fewest governance tensions identified within MPG compared with the other PRIs is in line with the stable evaluation mechanism with the fewest changes in dynamics. In contrast, will the more complex governance tensions match the more frequent changes in the evaluation mechanisms of HGF and CAS?

HGF

According to the case study, the *Programme oriented Funding (PoF) evaluation* of HGF has developed significantly in the past 15 years. Since the research activities of individual centres were strategically restructured under the umbrella of HGF, the uniform PoF approach aims for highly aggregated strategic research by

concentrating assets in multidisciplinary teams and by packaging the wide range of resources in centres. The **alignment of strategic goals** becomes the most important criterion in selecting programmes and topics to be funded. In particular, the three rounds of PoF evaluation have increasing consideration of the governance tensions found by this thesis, such as the centres' alignment versus autonomy, and their funding competition versus R&D cooperation.

The interviewees agreed that they gained experience and progresses in the past **three rounds of PoF evaluation**, as shown in Figure 6.2. The PoF I, at very early exploratory phase, aimed to support the full costs of the centres without any competition. It was more like a negotiation with the centres about which research and fields were to be supported in the long term. The impact of the evaluation results was limited. An important suggestion was proposed that the following PoF evaluation should provide more recommendations rather than requirements. The PoF II³¹ was a milestone: competition and cooperation, starting from very few cases, incrementally became visible at the level of programmes and topics involving participating centres. The centres officially started to work together and negotiate how to organise their cooperation and responsibilities. The PoF III, since 2013, shows two substantial changes. First, as the centres get to know more about each other, more interactions are required about drafting proposals, negotiating evaluation schedules, collective responses, and reactions to evaluation reports. In addition, the third PoF started a special programme, *Society, Energy, and Key Technologies*, where societal impact first became one of the core values at HGF.

The evaluation culture, the way people look at PoF—as interviewees said—has also changed along with the three rounds. The boundaries of disciplines and locations of the centres became fewer and the interaction between them increased. The acceptability of evaluation results by the interviewed scientists has grown, round-by-round, not at least because of the stable increase of programme funding. The centres have become more coherent and ready to discuss how to get additional synergies, rather than being competitive about funding, as explained by a high executive at HGF headquarters:

Now the competition is more on the level of honour and psychology of the powerful centres. That is why I think HGF really evolved

³¹ PoF II took place from 2009 to 2013 for four research fields and from 2010 to 2014 for the other two fields.

strategically more than other organisations. Our centres develop more and more together with each other. We had incompetency of the handling of the centres' strategy and PoF strategy in the very beginning of the PoF. It was very difficult but now it is under significant development.

CAS

CAS has experienced a series of governance reforms since 1997, stimulating more open and international evaluations. In particular, since 2012, *the One–Three–Five evaluation* has considered as a milestone in the institutionalisation processes. The **four stages of the CAS institute evaluation** have developed in the past two decades (Li et al., 2009); they are interpreted by this thesis as a formative way of institutionalisation of evaluation as a governance instrument. Some supplementary information can be found in Appendix C4.

Before 1993 there was no regulation about the frequency and indicators for institute evaluation, and no direct funding impact because the institute governance at that time was quite restricted by the planned economy in China. During the *first* stage from 1993 to 1998, evaluation started relying on quantitative outputs like publications and patents without consideration of impact. Then, the *second* stage was from 1998 to 2004, during the Knowledge Innovation Programme (KIP), when institute ranking by reviewers' scores appeared. Driven by lasting quantitative evaluations, the CAS's quantitative outputs (publications and patents) increased significantly at the end of 1990s and early of 2000s when new targets of research quality emerged in KIP and Innovation 2020. KIP, as a milestone of the governance, began to lead CAS R&D towards international standards (KIP Evaluation report, 2010). It was the turning point for institute evaluation to start to address qualitative evidence and impact assessment for governance decision-making. As Dai et al. (2014) show, this ranking directly impacted institutes' innovation funding, annual salaries of the directors, and other quota for staff positions and for applying for talent programmes.

Given the diversely developing institutes, according to Dai et al. (2014), the *third* stage was a comprehensive evaluation (2005–2010) with strong policy orientation and increased focus on institutes' innovation capability and creativity. Seven pilot institutes of the KIP-driven evaluation in 2007 introduced key qualitative indicators and peer-review-based benchmarks regarding different types of institutes. To

overcome the limits of purely quantitative indicators, a quality-oriented comprehensive evaluation methodology was promoted for improved institute evaluation, including past performance, current status and future plans of CAS institutes. This model was a pioneer, emphasising the different characteristics of research under different orientations and different research units. The innovation capacity of institutes became core objects monitored by the headquarters through a comprehensive indicators system referring to five aspects and 24 indicators, including self-status information and comparable data between institutes (Appendix C4). Qualitative evaluative methods based on expert judgement began to be used about the longer-term capability of institutes and their strategic direction, resources allocation, efficiency, and performance management. This system aimed to categorise all the institutes into the three groups (A, B, C). The results were linked with innovation funding—keeping the basic for C group, extra for B group, and additional extra for A group. This kind of evaluation exerted more pressure on the reputation of institutes and directors.

When CAS faces the challenge of fragmented R&D and inefficient investment as presented in sections 5.2 to 5.4, since 2012, the *Major Outcome Orientation* has been promoted depending on the *One–Three–Five evaluation* as the *fourth* stage. The *One–Three–Five evaluation* integrates the scattered research resources to improve the overarching mechanism at all levels by setting strategic plans and standards, replacing the previous fragmented research in short-term competitive projects. The development of the above four phases of evaluation institutionalisation at CAS has been driven by the requirements and interventions of governance, particularly on the strategic arrangement at L1 and funding adjustment at L2.

Because of being institutionalised, the evaluation mechanisms at the three PRIs are supposed to be accepted as mature and clearly defined instruments for governance uses. According to interviews, the ideal evaluation culture relies on the credibility and acceptability of evaluation results and their impacts. It is found that, so far, the acceptability of evaluation is more stable and positive in MPG and HGF than in CAS. However, the developing evaluation mechanism at CAS is becoming more strategic, open-minded, and international and brings about less mechanistic monitoring and more learning opportunities. The evaluation institutionalisation analysed in this section is closely related to the stakeholders at the three PRIs—their coordination

on evaluation procedures and acceptability of evaluation results—which is to be discussed in the next section.

6.3.2 Evaluation Facilitating Stakeholders' Participation and Interaction

As discussed in section 2.3, stakeholders' participation and interaction in evaluation activities are supposed to facilitate their negotiation of diverging interests. This section presents the stakeholders' participation and interaction observed within the evaluation mechanisms at MPG, HGF, and CAS. A common intention is found in normative evaluation procedures across the three PRIs—**increasingly involving multi-level stakeholders and providing opportunities for them to interact**. Based on the previous sections, stakeholders' participation is one of the key indicators to illustrate several key concepts of this thesis: *individual evaluation activities* (Table 6.1) and *evaluation mechanisms* (Figure 6.4), and closely related to many instruments within the *evaluation institutionalisation* (Table 6.2). All of these validate the element of *stakeholders' negotiation* in the overarching evaluation mechanisms for balancing governance tensions.

Across the three PRIs analysed, the evaluation mechanisms enrich the dialogues of the multiple stakeholders through *phases* of evaluation procedures. At the preparation stage, the *evaluation guidelines* are negotiated by the stakeholders and updated round by round thereafter. The guidelines not only distribute tasks and responsibilities to the multi-level stakeholders, they justify public investment and emphasise alignment of organisational missions and strategies. For example, the funding bodies at the national levels have increasing involvement at HGF and CAS in imposing evaluation objectives and management instructions. There are often several rounds of negotiation on such guideline-making and correction between decision-makers at L1 and L2. Moreover, external professionals in the scientific communities, the internal directors, and researchers at L2 and L3 increasingly engage in negotiating evaluation guidelines where they emphasise the selection of appropriate evaluators and evaluation criteria for research excellence. At both MPG and HGF, the Senate members from diverse backgrounds (politics, academia, business) broaden and strengthen strategic negotiation at L1. Comparatively, instruction-related negotiation between stakeholders mainly takes place only at L1 for CAS.

The following *evaluation procedures* are found to be similar at the three PRIs. At the stage of on-site inspection, various interactions are intentionally designed by

the stakeholders, where governance tension-related communication takes place. In particular, professional communication among experts using scientific and technical languages, in forms of presentations and workshops, are considered as rich resources to legitimate research investment and to answer evaluators' questions. After the primary evaluation results, the evaluated subjects have the chance to defend themselves and exchange feedback with other stakeholders cross the three organisational levels. Many bottom-up discussions take place during this phase, which may change the final evaluation results. In the German cases, reporting procedures experience more rounds of question-and-answer responses, including verification by the evaluated subjects and interpretation by the evaluators. The evaluated subjects also maintain some rights on whether or not to follow evaluation recommendations. For instance, one MPI may find the potential synergies from extended evaluation of a cluster not applicable to its circumstances and decide not to follow the recommendations. Using its third-party funding, one powerful HGF centre could continue the research that the PoF evaluation suggests it to abandon.

At the phase of evaluation results and uses, there is an issue of balancing between confidentiality of evaluation results and participation of external stakeholders; the issue is mitigated by the use of **various versions of evaluation results with different degrees of confidentiality**. What information should be transferred to which stakeholders, and to which degree: the original, detailed evaluation results from experts, the impact assessment of individual evaluated subjects, or summarised evaluative information across several subjects? This will be discussed in the following individual cases.

Overall, each case study has a distinct configuration of interactive patterns between particular stakeholders. Figure 6.7 shows MPG has a vertical shape with more bottom-up communication than the others. HGF exhibits a cluster type, shown in Figure 6.8, centred on the core programmes. The CAS situation is mixed, with the least clear patterns, but increasingly developing towards the aggregated mode centred on the *One–Three–Five aspects*, as shown by Figure 6.9. The following analyses will present the individual cases.

MPG

Stakeholders' participation and interaction is considered quite important, according to all the interviewees at MPG, to embrace the *evaluation culture*—the high level of acceptability and credibility of evaluation results made by the *best scientists* within particular fields. The most important target for evaluation organisers is to ensure that the existing evaluation mechanism involves the world leading experts. Figure 6.7 illustrates comprehensively the interaction patterns of the key stakeholders observed taking place within the evaluation mechanism across the three organisational levels.³² The arrows in Figure 6.7 represent particular communications and interactions taking place in evaluation procedures. The specific procedures and communicated contents are not specified in this section. Some of them have been presented in section 6.1 and some can be found in Appendix A4.

What is emphasised here is the different types of evaluation results and the varying degrees of accessibility for various stakeholders. *Ex-ante* evaluation about appointment of the institute director takes place only among heads within MPG. Results of both the extended evaluation and strategic bibliometrics only serve L1 deciders, such as the headquarters, for particular policy-making tasks. *Ex-post* institute evaluations yield several reports—the most detailed version goes directly to the MPG president, an interpreted version goes back to institutes for clarification and defence, and an updated version, after rounds of feedback, goes to the Executive Committee. The Senate only gets summary results on the general situations of institutes that are readily understandable and in which the Senate can easily intervene. The reason why the original evaluation results are kept confidential was explained by a high executive in the MPG headquarters:

SAB members trust that their evaluative findings would not be published, and, particularly, not forwarded to governments, but directly to the president who decides which part is contained in official summative reports. Therefore, very frank and open recommendations could be made. Otherwise evaluation reports would lack quality.

³² Due to the limited space, some stakeholder labels should refer to Table 4.2 in section 4.4. For example, A2 refers to Executive Committee.

Figure 6.7 Interaction model of MPG's evaluation mechanism

Source: the author, based on a report of MPG evaluation procedures (Schruff, 2012)

Most internal stakeholders at L2 and L3 are only able to access the interpreted versions of evaluation results. The compromise of doing this is that evaluation guidelines and procedures are required to be equally transparent for all the organisational levels including keywords, defining funding areas, selection procedures of evaluators, assessment criteria, etc. The most important guidelines—SAB rules—are transparent, available online, and claimed to be strictly implemented. Therefore, the stakeholders at each level are clear about their jobs. Such a **balance between clear transparency of guidelines and confidentiality of evaluation results** aims not only to improve the efficiency and objectivity of evaluation implementation, but also to mitigate the governance tensions identified in section 5.2—the external justification versus internal governance of heterogeneous research.

Such a vertical interaction style at MPG is more bottom-up than at HGF and CAS because its core aim of research excellence is mostly achieved by individual scientists at L3 whose high level of scientific freedom and trust are more protected rather than regulated by the evaluation mechanism. In contrast, HGF and CAS face the stronger tension between research management instructions and scientific freedom at L3, such as the ability for individual researchers to pursue their own initiatives, identified in section 5.4. How will they facilitate the stakeholders' participation and negotiation in the evaluation procedures?

HGF

HGF has the closest connection with politics in the German research system: national strategic research fields are defined jointly by the government and HGF. The government also requires responses on certain political expectations like internationalisation of research and education and recruiting talent in particular areas. Such close interaction with political stakeholders is also reflected by its negotiation on the **Strategic Guidelines** of PoF evaluation, including the key scientific areas, structure of the fields, general review objective and criteria, and Five-Year budgetary framework. Such negotiations reflect the public justification of HGF and consider the internal governance on strategy and funding. Therefore, the participation of and interaction between the external and internal stakeholders on *Strategic Guidelines* help to balance their conflicting expectations of HGF as identified in section 5.1.

Figure 6.8 Interaction model of HGF's evaluation mechanism

Source: the author, based on the interview findings, 2013

Research programmes are the core connecting the stakeholders, as shown in Figure 6.8. As discussed, PoF implementation and evaluation aggregate the individual centres as well as their internal departments of Programme Planning and Management, which are particularly responsible for coordinating the multiple stakeholders across different programmes. The other very important guideline—the **Position Papers**—focus on the definitions of the major scientific challenges and goals in each field and of research activities cross the programmes (Appendix B4). The Position Paper, guiding the specific R&D activities in each field, is constantly discussed and updated in relation to research fields involving several programmes across the centres. Such discussion is organised by the Senate on dialogue platforms between representatives from science, politics, and industry and from internal centres. The negotiations focus on how to follow and interpret the overarching policy goals proposed in Strategic Guidelines into specific R&D targets of each field defined in Position Papers. Such negotiations help to balance the governance tensions at HGF on the aspect of strategy identified in section 5.2—

conflicting goals from interpretation of overarching organisational strategies by member centres.

With these two key guidelines, the following evaluation procedures are very similar to those of MPG's. At the stage of evaluation results, the process from evaluation reports of individual programmes to the final comprehensive PoF decisions and recommendations of all the programmes involves multiple interactions—who contributes to which reports, who gets access to which reports—and the confidentiality and feedback on various versions of the reports (Appendix B4). Like MPG, only top-level managers have access to the original PoF evaluation reports proposed by review panels, like the two Boards (SAB and Supervisory Board), and directors. They discuss the detailed recommendations and respond with their feedback. The headquarters then collects reports on individual programmes as well as the feedback from the centres and writes a first draft of the funding recommendations and requirements for each programme. This is a collective work in close consultation with the multiple stakeholders centred on the programmes—Senate Commission, chairpersons of review panels, reviewers, and coordinators of research fields, programme spokespersons, and the Boards of the participating centres. Then the Senate Commission makes a comparative strategic evaluation of different programmes within one field based on the individual reports, and prepares the comprehensive financing proposal to the funding bodies that decide on the funding of the programmes and centres for the next period. The Senate also discusses the final decisions and gives its opinions. Therefore, the negotiations in the whole process, from evaluation results of individual programmes to the final PoF decisions and recommendations of all the programmes, contribute to balance the funding competition between programmes within each field and the participatory centres in each programme.

To sum up, the interaction style in the PoF evaluations at HGF is in line with its *programme-centred* governance emphasising more strategic alignment with defined objectives than research quality. Top-down requirements, mostly in strategic guidelines, and bottom-up scientific responses are observed interweaving. As discussed, this style aims to provide reliable and legitimate evidence for PoF decision-making involving several governance tensions. The midterm reviews of each centre are not discussed here because they focus on scientific evaluation of research quality, the results of which are also included in the PoF evaluation.

CAS

The stakeholder interactions at CAS, both internally and externally, are neither as frequent nor as normalised as those at MPG and HGF. Participation and interaction with and between the stakeholders are observed as in-house and random in the past evaluation activities. The fragmented funding-oriented projects as the reason of several governance tensions, as presented in sections 5.2 and 5.4, also result in a situation where stakeholders negotiate such tensions in an isolated manner. Compared with the German scientists, the interviewed Chinese scientists are more shy and conservative to share opinions on interaction with stakeholders. However, in recent years, the following three areas of progress were observed, which result in the emerging of a clearer interaction pattern towards the aggregated *One–Three–Five aspects* to be evaluated, as Figure 6.9 shows.

First, CAS communicates with the State Council through national policy-making and the related evaluation reports. The KIP was a breakthrough in relation to the translation of political targets into scientific goals and practices. A few pilots of KIP were assessed in midterm and *ex-post* evaluation; related reports were pioneering in the way that they publicly announced the effects of governance reforms of CAS. In the 2012 National S&T Innovation Conference, CAS as the pilot explorer proposed the Major Outcome Orientation and *One–Three–Five evaluation* as scientific responses to political requirements. Particularly after Pioneer Action Plan (PAP), CAS has become more capable of providing evidence to national policy-making through consideration of classified research orientations. The evaluation of Major Outcome Orientation is expected, by the author, to bring more waves of political and scientific negotiations.

Interactions between the State Council and CAS regularly take place in national political conferences through annual performance reports as *ex-post* evaluation. The State Council and CAS also negotiate new policies as *ex-ante* interventions to CAS plans. Yet, the intermediate monitoring is mainly done by financial auditing procedures. The political targets are implemented as centralised guidelines strictly in a top-down way from the State Council to the CAS headquarters, then to research institutes.

Second, interaction between CAS and external stakeholders usually takes the form of drafting and assessing national large-scale and strategic programmes. The *Chief Expertise Review Panel*, as the highest national-level advisory board, includes

scientists and academicians from MOST, CAS, universities, and some enterprises. Then the programmes are divided into projects with these experts as promoters for the involving institutes. Such high-level interaction between politics, the academy, and industry used to occur only in a very limited number of national events, but now is becoming more frequent because of the increase in strategic programmes. In the *ex-ante* evaluation of PAP, as presented in section 4.2, the *Chief Expertise* experts and especially established Advisory Committee with representatives from the ministries were consulted to obtain diverse suggestions. These negotiations are considered innovative and contribute significantly to the CAS institutional reforms.

Figure 6.9 Interaction model of CAS's evaluation mechanism

Source: the author, based on interview findings, 2013

Furthermore, some other unstructured types of interactions, as the dotted line in Figure 6.9 shows, are increasingly found and have great potential to contribute to

mitigate the governance tensions. For instance, interaction between reorganised Bureaus within the CAS headquarters is taking place because of their increasingly interdisciplinary work (Appendix C3). Furthermore, collaboration between research groups of different institutes is promoted by the new research clusters in PAP and by the increasing requirement—the aggregated Major Outcome. Within the processes of defining and assessing the Major Outcome and *One–Three–Five aspects* for each institute, negotiations between internal research groups, between the institute and the headquarters, and between the institute and international experts are all increasing in terms of both numbers and frequency, and people's acceptance. Those negotiations gradually go beyond R&D improvement but seek to contribute to the reformed governance of CAS about strategy and funding. These reforms that aim to mitigate the governance tensions at CAS, as discussed repeatedly, increasingly rely on the evaluative evidence. Interviewees observed that the expected *evaluation culture* at CAS—credibility and acceptability of evaluation impact towards organisational governance—was emerging at a slow pace.

Across the three case studies, besides the above findings on stakeholder interaction in formal procedures, the interviews show some informal practices that are also considered important. In fact, people take and learn from both formal and informal practices. Some unplanned interaction of stakeholders takes place during on-site evaluation activities. Scientific negotiations facilitate better understandings between evaluators and the evaluated subjects, sometimes with unintended products such as creative ideas beyond evaluation settings. Negotiations on governance aspects may lead to changes or additionality of stakeholders' behaviours that could be further considered and taken advantage of by people or the institutions. Personal networking was not looked at in-depth in this study but that could be another possible outcome. The two approaches of the evaluation mechanisms at MPG, HGF, and CAS—*evaluation institutionalisation* and *stakeholders' negotiation* are interpreted by this thesis as their balancing processes to reflect and mitigate the governance tensions. These findings contribute to answer how the evaluation mechanisms facilitate organisational governance. Based on that, the next section will present further evidence on how this works at the three PRIs in a higher and more integrated way—that is, in learning.

6.4 Evaluation Mechanisms Facilitating Organisational Learning

Science itself is changing, developing and learning all the time, as well as the related governance and evaluation of scientific work. Based on the evidence from MPG, HGF, and CAS, this section validates the proposition discussed in section 2.4 that evaluation mechanisms facilitate organisational learning in two ways. The first is that the evaluation mechanisms provide three types of knowledge bases for governance. The second is that evaluation mechanisms' learning from stakeholders' interactions and negotiations facilitates organisational learning.

6.4.1 Knowledge Bases for Governance

As the case studies show, the evaluation findings explicitly contribute to the organisational governance through different types of knowledge bases. The three types of knowledge bases, discussed in section 2.4.1, are confirmed based on the following evidence.

The *analytic knowledge bases* are the largest of the three types produced from evaluation and used for governance at all the three PRIs. This type of knowledge is generated from results and recommendations of professional R&D evaluations relying on experts' analyses and judgement. The goals of such R&D evaluations are quite objective and focus on the improvement of R&D qualities within the scientific communities. The recommendations are usually focused on the evaluated subjects, rather than on governance in general. For the three PRIs, such analytic knowledge bases can impact on setting scientific goals of individual evaluated subjects in a specific discipline or field. This type of knowledge base is used more by the evaluated subjects themselves than by the whole institutions, in other words, like the institutes of MPG and CAS, the HGF centres at L2 and all the scientists involved at L3. Additionally, evaluation methods are constantly updated by learning from interactive workshops and conferences within scientific communities, which contribute to the analytical knowledge creation.

The *synthetic knowledge bases* target potential synergies among several evaluated subjects for compound effects of the higher-level decision-making, usually for the whole institution. Such knowledge bases are generated from comparative evaluations across individual subjects, such as extended evaluation at MPG across several institutes in similar fields, and programme evaluation at HGF across several centres. Normally, such synthetic knowledge has to be generated based on the

analytical knowledge when evaluating individual subjects. For instance, the extended evaluation at MPG is based on each institute's evaluation results; the strategic decisions on funding programmes at HGF depend on the centres' own scientific midterm reviews. For CAS, the integrated *One–Three–Five aspects* of each institute to be evaluated are also internally assessed as individual research topics.

This type of synthetic knowledge is found particularly useful for balancing competition and collaboration between the units at L2, especially for the funding competition between the centres at HGF and between the institutes at CAS. In these two cases, richer experiences of the comparative evaluations and more synthetic knowledge are gained. Generally, interdisciplinary research asking for more collaboration across multiple disciplines, fields and organisational levels is more intended to create synthetic knowledge from the evaluation. In sum, decision makers coping with the complex governance problems can receive synthetic evaluative information integrated from reliable evaluation evidence, such as funding HGF's programmes across fields and centres.

The *symbolic knowledge bases* are supposed to be the verification of the authority, reputation, and prestige of the evaluated subjects (such as research topics and researchers) by top professionals. Such knowledge is usually used to strengthen the credibility and reliability of the above two types of knowledge. In this sense, the higher reputation the evaluators have, the more symbolic knowledge bases are produced. For example, all the evaluations at MPG rely on such symbolic knowledge bases for governance use, which is more obvious there than in the other two cases. The key governance decisions at MPG—establishment of new institutes and appointment of the institute directors—rely much more on the symbolic knowledge bases than at HGF and CAS because the evaluators at MPG are supposed to be the *best scientists* within the particular fields.

Additionally, such knowledge bases create more value if the credibility of reviewers justifies public investment in some unique research topics worldwide. According to interviews, such one-and-only topics could be relatively unpopular fundamental research fields with limited researchers interested in the fields, as Arnold (2004, p. 9) calls it, “‘useless’ research fields (astronomy? particle physics?)”, or strategic research fields in unique or complex contexts.

Before evaluation became a standardised exercise in CAS, internal seniority rather than external expertise contributed to symbolic knowledge bases for a long time.

Since then the increasing international evaluations at CAS have led to growing symbolic bases, while the evaluation culture has become more open, understanding, and acceptable, instead of being quite closed.

The three knowledge bases produced from the evaluation mechanisms are found to contribute, with their own foci, to different types of governance decisions at the three PRIs. As a result of the growing complexity and diversity of knowledge creation and application, governance decision-making at the three PRIs increasingly uses evaluation mechanisms to strategically integrate various knowledge bases. This is a trend and also a long-term learning process.

6.4.2 Organisational Learning Fostered by Evaluation Mechanisms' Learning

The balancing processes of the evaluation mechanisms, as discussed in section 6.3, also include learning processes of the evaluation mechanisms. The empirical evidence of the three PRIs helps to extend the theoretical definition of learning evaluation, as discussed in section 2.4.2. Evaluation activities taking place in the three PRIs facilitate continuous improvement by learning from R&D-related evaluative recommendations and their uses for governance purposes. According to section 6.3.1, the institutionalisation of the evaluation mechanisms over time can be seen as a systemic feature by the expansion of governance concerns. In this sense, the institutionalisation of evaluation mechanisms can be considered as resulting from *learning through evaluation experiences*, eventually making the overarching evaluation mechanism responsive to the governance requirements, such as the guideline updates in MPG and HGF. Furthermore, the evaluation mechanisms of the three PRIs are found to respond with growing capacity to various stakeholders and reflect stakeholders' tensions, as discussed in section 6.3.2. Additionally, the three PRIs also learn from international experts. For instance, CAS observed and learnt from multiple international evaluation experiences, by drawing on the published documents and inviting global experts, for planning and implementing its current *One–Three–Five evaluation*.

How does the above learning within the evaluation mechanisms facilitate organisational learning? Key approaches as well as resources found at the three PRIs are about stakeholder interaction and negotiation across the three organisational levels (L1, L2, L3). The evaluation mechanisms involving multiple stakeholders' participation and interaction, as presented in section 6.3.2, gradually

drive the PRIs to become increasingly **responsive to the diverse stakeholders as well as their governance tensions**. For instance, at the planning stage, the evaluation organisers learn from a wide range of stakeholders who negotiate the evaluation guidelines from their diverse perspectives and international experiences. Usually the evaluation objectives are broadened with more consideration for stakeholders' conflicting expectations. Then the on-site evaluation activities bring about learning opportunities for both researchers on scientific learning and governors on how to find credible ways of mitigating governance tensions. During the evaluation impact phase, organisational learning takes place along with the transfer and negotiation of the stakeholders' feedback, which makes evaluation results and recommendations more accurate, reliable, and effective to be taken advantage of by the organisational governance.

Combining sections 6.4.1 and 6.4.2, Figure 6.10 visualises the process whereby the evaluation mechanisms' learning facilitate organisational learning. The driver—on the left side—is the institutionalisation of evaluation mechanisms to become increasingly systemic, at larger scales, and more responsible to stakeholders. The stakeholders across the three organisational levels learn from two approaches. The first is that explicit evaluation findings are utilised as the analytical, synthetic, and symbolic knowledge bases for governance decision-making. The second approach of organisational learning takes place in the process of stakeholders' interactions and negotiations across the three organisational levels (L1, L2, L3), as shown by the three gears in Figure 6.10. In fact, the two ways of learning, from evaluation findings and processes, supplement each other.

Furthermore, the organisational learning of the three PRIs is considered as a part of their evolving organisational governance to balance the stakeholders' tensions. Based on the coordination of the stakeholders at the three levels, just like the co-movement of the three gears, the learning activities in the evaluation mechanisms are observed as transferring, in a bottom-up way, from scientists at L3 to institutes at L2, then to the whole institutions at L1. The core reason found for this is that scientists who conduct R&D directly and individually learn from R&D evaluations as well as recommendations. The example is the analytical knowledge bases produced from R&D evaluation for scientific decision-making. The organisational learning at L2 and L1 takes advantage of the more integrated and synthetic knowledge bases to mitigate the *governance tensions* across the three levels as identified in Chapter 5, just like the friction between the three spinning gears that needs to be overcome.

In this sense, the evaluation mechanisms provide not only the drivers, but also the lubricants to the organisational learning.

Figure 6.10 Evaluation mechanisms facilitate organisational learning

Source: the author, 2015

Therefore, the organisational learning of the three PRIs is found to frame appropriate uses of the evaluation mechanisms, both findings and processes, as instruments to achieve better organisational governance. Furthermore, the organisational culture of the three PRIs is fostered through learning from the evaluation mechanisms. The ideal governance effect after learning will be discussed in the next section.

6.5 The Balanced Organisational Governance

Consequently, although in significantly different contexts, the three evaluation mechanisms at MPG, HGF, and CAS are found to reflect and balance their key governance conflicts to varying degrees. The balancing processes within the evaluation mechanisms are expected, by both the author and the interviewees, to result in **desired governance results** at each of the three organisational levels, as

shown in Table 6.3. Responding to Table 5.1 showing the governance tensions level by level at the three PRIs, Table 6.3 illuminates the expected governance results with mitigated tensions. The commonalities and differences of the three PRIs are shown by the shared, unique, and empty cells in Table 6.3. In fact, some expectations, by the PRIs themselves and the author, are not fully realised yet and could be validated only if evaluation activities indeed lead to improved governance some years later. Some expectations are almost achieved while some are in the early stages because governance conflicts always change along with the stakeholders' responses in the balancing processes.

The balanced results	The PRIs studied		
	MPG	CAS	HGF
Governance equilibrium maintenance at L1	Tailored scientific governance and highly standardised administrative governance; Managing bottom-up scientific influences and their strategic uses		
	Administrative control from the headquarters		
		Priorities are made for funding battles	
Autonomous development at L2	Scientific autonomy is improved by strategic alignment		
		Institutes and scientists are more empowered by a higher proportion of block funding	Centres and scientists are more empowered by stable increases in block funding and flexible budget
Freedom and trust at L3			
	Rigorous evaluation protects scientific freedom and trust		

Table 6.3 The balanced results of governance of the three PRIs

Source: the author, 2014

It is worth mentioning that in the interviews, the impression of the concept of *balance* was viewed much more positively in the Chinese than German contexts. The German understanding of balance is neutral, similar to the moderation medium (German: *Mäßigung*) of struggles and negotiations proposed by Kuhlmann (1999). The understanding of balance in Chinese society follows the idea of a *Harmonious Society* which was politically raised and was emphasised in the past decades as a primary strategic goal in the eleventh Five-Year Plan to lead public policies. *Harmonious Society* calls for democracy, justice, trust, friendship, and well-ordered relationships between nature and human beings. This concept of *harmony* is derived from traditional Chinese Confucianism, which believed in the unity of

ideas, behaviours of people, society, and state in an ideal condition which is peaceful, harmonious and happy for everyone. This ideal situation was proposed as a prioritised political target aimed at solving the tension between the high speed of economic development and the slower advancement of Chinese society and culture. The ideal condition—a peaceful and balanced life after overcoming those identified or predicted conflicts—has also been expected by the Chinese scientists. The following analyses will present the expected balanced governance level by level.

6.5.1 Governance Equilibrium Maintenance at L1

Reading Table 6.3, across the three PRIs facing multiple stakeholders and their conflicting interests and power, the overall organisational governance at L1 all aims to be maintained in equilibrium. To achieve the *governance equilibrium*, the decision-makers and supervisory bodies at L1 have to overcome the key governance tension at L1—external justification of public funding and internal governance of heterogeneous research, as discussed in sections 5.1 and 5.2. This thesis expects two specific balances to overcome this typical tension at L1.

The first balance is to empower **tailored scientific governance** for the heterogeneous R&D units at L2 and meanwhile to keep them under **highly standardised top-down administrative governance**. Tailored scientific governance relies on the credible evaluation evidence and knowledge bases generated from the strictly implemented R&D evaluation procedures in relation to the individual subjects. The overarching evaluation mechanisms are institutionalised at the three PRIs to standardise the administrations of the heterogeneous R&D units at L2. Such a balance between *tailored* and *standardised* ways at L1 helps with the challenge of governing the heterogeneous R&D units at L2.

For example, at MPG, the most frequent institute evaluation—which aims at tailored scientific governance at each institute—is mainly taken by the institutes' own SAB members. Meanwhile, evaluation guidelines concerning evaluator selection and criteria are standardised and continuously updated to ensure the guidelines meet the governance requirements of the whole institution. Furthermore, there are regular announcement of summarised evaluation results to justify the public trust.

The second balance is between the top-down requirement of strategic alignment and the bottom-up scientific responses. The former is accomplished by strategic evaluation guidelines that address the requirements of external national funders, such as the Strategic Guidelines at HGF and the Major Outcome Orientation at CAS. However, detailed evaluation results and recommendations for internal governance are still highly confidential. For both HGF and CAS, this type of balance is taken by the centralised strategies and decentralised R&D activities. Both PRIs justify the latter, as well as the developmental autonomy, by the self-assessment of the units at L2. The more satisfactorily the units behave in both strategic alignment and scientific reviews in one evaluation round, the more resources they receive for their autonomy for the next round. For MPG, its individually developing research, low financial competition and higher level of scientists' freedom are limited by centralised administrations of the institutes. Such limits are the same for CAS in terms of administrative control of the institutes by the headquarters.

Additionally, the internal governance of heterogeneous R&D depends on both scientific assessment of individual R&D units and of clusters, like MPG institutes and HGF programmes. The cluster evaluations pay close attention to and make recommendations on institutional factors (such as institutional change across the individual units) for their potential synergies. They contribute to decisions on resources allocation within and between the evaluated units, as well as between exploitation of current excellent R&D and exploration of potential R&D. Typical examples are the programmes at HGF, and the aggregated Major Outcome for CAS institutes. At CAS, a new systemic evaluation integrating the previous isolated ones was created. The *One–Three–Five evaluation* is a significant step in categorising complex governance of a huge number of institutes and facilitates stability and predictability of the whole CAS system (Luo et al., 2015). In this way, more centralised scientific evidence is produced to depersonalise and depoliticize decision-making on resources allocation at L1.

Therefore, a delicate balance is ideally pursued between public justification and internal governance of heterogeneous research at the three PRIs. Overall, their evaluation recommendations at L1 and the general institutional dynamics always go hand-in-hand.

6.5.2 Autonomous Development of Institutes at L2

A common target observed across the three PRIs—the autonomous development of the units at L2—is promoted. As discussed in sections 5.2 and 5.3, the governance tensions at L2—on strategic alignment and funding competition—mainly result from the varying degrees of the autonomy of the units, whether they are legally independent or not. There is always a **balance between the strategic alignment and the autonomous development** of the units at L2, for MPG and CAS institutes, and HGF centres.

Across the three cases, the *scientific autonomy* of the units at L2 is generally at a high level because their mainstreamed R&D is legitimised by the evaluations, which strengthens their autonomy to conduct other R&D funded by third parties. For example, MPG institutes, while being legally dependent on the headquarters' governance decisions, have the highest level of scientific freedom and trust. This is interpreted as a kind of autonomy that is also frequently monitored by high-level international experts. HGF centres and CAS institutes experience less frequent but more strategic evaluations. Meanwhile, they both have empowered reviews for their internal governance. Therefore, legally independent HGF centres and CAS institutes balance overarching policies and autonomous development by following strategic alignment and organising their own scientific reviews. HGF centres and CAS institutes take and learn from both types which supplement each other.

The MPG institutes, the legally independent centres at HGF, and institutes at CAS all have the financial autonomy of internal resources allocation. The higher the proportion of block funding, the more empowered governance the units enjoy. MPG and HGF already have a high level of block funding ensured by their most important evaluations: *ex-ante* institute evaluation at MPG, and PoF evaluation at HGF. For them, a small proportion of third-party projects together with the headquarters' flexible funding slightly supplement the core finance. With a shortage of block funding, as discussed in section 5.2, the Major Outcome Orientation at CAS actively promotes the aggregated *One–Three–Five aspects* of the institutes with a higher level of institutional block funding by around 30%. In this way, some previous resources on multiple research projects could be aggregated for focused developments. Likewise, more space is allowed for other R&D beyond *One–Three–Five aspects* funded externally. For CAS, on the condition that the Major Outcome is guaranteed by *One–Three–Five evaluation*, there are

many open policies to attract academic and industrial collaborators, both domestically and internationally. Their actual implementation still depends on the institutes' own scientific communities and scientists' networking, which is related to the discussion in the next section.

6.5.3 Protection of Freedom and Trust for Scientists at L3

Trust in science depends on the trustworthiness and reliability of the evaluated R&D and scientists, which is closely related to scientists' freedom and trust. As proposed in section 6.3.1, all the three PRIs are cultivating a culture where the credibility and acceptability of evaluation results protect scientific freedom and trust of the scientists at L3. The assumption proposed in section 2.1.1—the higher proportion of block funding, the more discretion researchers have over the selection of research topics and approaches and the more time they have to pursue ambitious intellectual goals—is confirmed for the three PRIs. Some interviewed researchers do not like evaluation administrations but all the interviewed researchers welcome critical R&D recommendations.

A mature organisational culture is found at MPG in that **evaluation promotes mutual trust** between public funders and internal scientists, and also between the headquarters and institutes. Their consensus is that the high level of scientific freedom and trust at MPG depends on the consistently implemented evaluation procedures of the past 45 years. Within the three PRIs studied, the MPG scientists have the highest level of scientific freedom and trust, which is not only because of fewer management instructions on research themes and topics, but also because of the culture—evaluation is considered as a two-year routine of assurance and further improvement. Evaluated researchers do not feel monitored by their outstanding peers, but instead are motivated to interact with them. MPG scientists always believe they will receive excellent feedback and recommendations from evaluations. The MPG headquarters believes that *ex-ante* evaluation clearly helps them ascertain whether a particular leader has a high potential to construct the best research in a field and matching the great reputation of MPG. The trustworthiness of excellent scientists and the person-centred funding framework of MPG are justified and guaranteed by its frequent and rigorous evaluations.

At HGF, when complying with policy guidelines of research programmes, the scientists still enjoy the **freedom to interpret the overarching policies** scientifically by defining their own research topics. Its evaluation culture is characterised by the

belief that scientific work proposed in a coordinated way by participating centres for the PoF evaluation will eventually lead to increasing budgets for the next five years. The preparation and coordination should follow closely the Strategic Guidelines and Position Papers which are formulated in a democratic way. HGF scientists have access to the pre-discussion of those guidelines and then stick to them strictly. Additionally, the 20% flexible budget of PoF without thematic requirements represents the scientific trust. This culture is cultivated through the previous three rounds of PoF evaluation.

Scientists' freedom and trust were discussed more often and highlighted by the German interviewees than the Chinese in this study. In the two German PRIs, the evaluation mechanisms aim to verify and improve the value of scientific work and scientists' trustworthiness. The concept of evaluation was understood by many Chinese interviewees as receiving and producing impact for financial audit and supervision which went against trust. The consensus among the three PRIs was that scientists should be allowed to have more freedom and trust in well-regulated PRIs. However, such flexibility may not necessarily be better. The more freedom the rules give to scientists, the more strictly and fairly the related punishments are implemented if the scientists break the rules. In Fukuyama's study of society and economy (1995), Germany generally shows a high level of trust within its organisations because of well-established regulations and a high level of technical skills of employees. German consensus would believe that well-skilled researchers in the trustful environment with stable institutional funding make great efforts and concentrate on their work which leads to excellent outcomes. However, many interviewed Chinese scientists, who are much less well trusted, are just given basic resources to push them to earn more by themselves through competitive research projects. A widely accepted hypothesis by the Chinese scientists is that people become greedy and lazy once they get rich resources too early or too easily.

As discussed in section 5.4, CAS scientists have the least freedom because of the overwhelming task of acquiring competitive project funding. Scientists look forward to higher block funding after the *One–Three–Five evaluation* to help them to get rid of the financial burdens and also to embrace more research flexibility and freedom. Chinese scientists are now experiencing fewer monitoring but more learning opportunities in more strategic, open-minded and international evaluations. Their freedom and trust are *increasingly protected* to promise the continuity of *One–*

Three–Five aspects in long term. It is a big step for Chinese scientists to accept evaluation as an effective instrument to boost scientific work.

To sum up the three cases, the scientists' trust and freedom require protection to ensure the continuity of their creativity, productivity, and commitment. Rigorous evaluation does not mean no freedom and trust. Good evaluation results verify scientists' creativity and productivity, and contribute to the excellent reputation of the institutes and the whole institutions, which also improves scientists' trustworthiness. The interaction between scientists and other stakeholders in evaluation activities boost the credibility of evaluation results and empower the evaluated scientists to be confident in the face of critical reviews. To selectively publish different versions of evaluation reports also protects research itself as well as scientific freedom.

To conclude this chapter, empirical evidence is provided to answer the two sub-questions proposed in the enriched conceptual framework in section 4.5:

- ② How do the evaluation mechanisms of the three cases reflect and mitigate governance tensions concerning the sources identified—governance aspects and stakeholders?
- ③ What are the balanced organisational governances for the three cases?

The analyses of the three PRIs present, individually and comparatively, the balancing processes of the three evaluation mechanisms—responding to the governance tensions in their institutionalisation and facilitating stakeholders' interaction and negotiation. These balancing processes of evaluation mechanisms facilitate the organisational learning across the three levels (L1, L2, L3). In addition, the three differentiated knowledge bases (analytical, synthetic, symbolic) are generated from evaluation findings for governance decision-making. Consequently, there are desired balanced governance results for the three cases at the three organisational levels; some have been realised and some are interpreted as the expectations. The empirical findings in this chapter will be combined with those in previous chapters to produce integrated conclusions and implications in the next chapter.

7 Conclusions and Recommendations with the Completed Conceptual Framework

Life is like riding a bicycle. To keep your balance, you must keep moving.

---Albert Einstein

In this final chapter we will review the overarching research question this thesis addressed: ***How do evaluation mechanisms of publicly funded research institutions (PRIs) play a balancing role and help to mitigate tensions in organisational governance?*** To answer it, the double causality between *organisational governance* and *evaluation mechanisms* has been explored through an abductive process of establishing the analytical instrument—the basic, enriched, and final conceptual frameworks. The final framework, to be discussed in section 7.1.5, is completed by exploring the four sub-questions, proposed in the enriched framework (section 4.5), in the case studies in Chapters 5 and 6:

- ① What are the typical governance tensions of the three PRIs concerning both governance aspects (strategy, funding, operation) and multi-level stakeholders (L1, L2, L3)?
- ② How do the evaluation mechanisms of the three cases reflect and mitigate governance tensions concerning the sources identified—governance aspects and stakeholders?
- ③ What are the balanced organisational governance for the three cases?
- ④ Do the balancing roles of evaluation mechanisms of the three PRIs perform differently in their different contexts, and if so, is there a generalisable principle?

The key proposition of this thesis is that evaluation mechanisms can help to reflect and mitigate governance tensions and facilitate organisational learning at PRIs; the two key approaches are institutionalising evaluation practices and facilitating stakeholders' interactions and negotiations. Centred on this proposition, section 7.1 presents key results of this thesis. The key findings of the case studies of Max Planck Society (MPG), Helmholtz Association (HGF), and the Chinese Academy of Sciences (CAS) are summarised in sections 7.1.1 to 7.1.4 to answer the four sub-questions. Based on that, the empirically refined and completed conceptual

framework will be presented in section 7.1.5. The theoretical contributions are summed up in section 7.2. Implications for institutional practices and public policy on research evaluation are identified in section 7.3. Limitations of the research done are noted and ideas for future research are proposed at the end of this chapter.

7.1 Key Results

An interpretive method (Yanow, 1998) was used in this thesis to translate and integrate the findings of the three independent cases into comparative analyses and generalised arguments, which will be presented in the following sections.

7.1.1 What are the Typical Governance Tensions?

Governance tensions were categorised into national (or system) and organisational levels in section 2.2. At national levels, both Germany and China have embraced significant investment and growth in research and development (R&D) activities with their own competitive advantages as well as challenges. An ongoing discussion was observed in Germany on how to improve further the national R&D landscape in relation to tensions in funding allocation between universities and non-university PRIs, and among various PRIs. It was found that China lacks a well-categorised R&D system; the public research resources were inefficiently used because of the mixed research orientations (fundamental, applied, strategic) of the national key R&D contributors. MPG, HGF, and CAS were found to have roles and challenges in their respective national R&D landscapes. Yet, regarding the national-level tensions discussed by the literature in section 2.2.1, little explicit evidence was found from the three PRIs' evaluations that responds to the national-level tensions. Implicitly, stakeholders' negotiation at L1 helps to consider and prevent the discussed national-level tensions in organisational practices. One key reason could be that MPG, HGF, and CAS are not required to be part of national evaluation systems. The three PRIs design and use their own evaluation mechanisms mostly for their internal governance, which is the main finding of this thesis and will be presented in the following sections. Based on the key findings in the three PRIs, this thesis also proposed extended reflections on public policies at national levels, which will be discussed in section 7.3.2.

At organisational levels, 15–20 years' dynamics of organisational governance as well as evaluation practices of the three PRIs (MPG, HGF, CAS) were reviewed. First,

eleven observable dimensions were found to specify the abstract concept of *organisational governance* into institutional characteristics of national research systems and internal research behaviour and decisions (see Table 4.1). The three PRIs performed differently on the shared dimensions but some underlying common principles were found. Importantly, all of their different characteristics are in line with their corresponding organisational missions (e.g. to expand boundaries of knowledge or to contribute to national development) and research orientations. For example, MPG's mission to do *the best fundamental research worldwide* decides its person-centred governance of research institutes and a high level of block funding (85%).

The three-level organisational structure was confirmed at the three PRIs: central decision-making and supervisory body at the macro-level (L1), research institutes at the meso-level (L2), and individual researchers at the micro-level (L3). *Fourteen groups of key stakeholders* of MPG, HGF, and CAS were categorised that can be shared by PRIs in general, such as internal scientific advisory board and external public funding bodies (see Table 4.2). The stakeholder groups were found to have diverging interests in the governance aspects and different power at particular organisational levels, which were considered by this thesis as the sources of governance tensions.

Based on that, **similar types of governance tensions** were found in the three cases, as shown in the three-by-three matrix of Table 5.1 about governance aspects (strategy, funding, operation) and organisational levels (L1, L2, L3). Some types have two main conflicting components, as shown in Figure 5.1, such as a quantitative balance between proportions of institutional block funding and competitive project funding. Other types of tension are qualitative, such as the conflicting expectations of internal and external stakeholders (section 5.1).

At L1, the organisational strategy of PRIs needs balanced consideration of both external justifications by completing publicly defined missions and internal governance of heterogeneous R&D activities in different research disciplines, fields, and institutes. A common conflict found in the three PRIs is that their overarching policies at L1 are not one-size-fits-all for the large number of units at L2. Furthermore, the heterogeneity of the units at L2—institutes at MPG and CAS, and centres and research programmes at HGF—worsens this type of tension. At L2, the most important balances to be achieved are between competition and

collaboration with other units, and between aligning organisational strategies and those strategies developing autonomously. At L3, scientists' freedom and trust, such as the ability for individual researchers to pursue their own initiatives, can be threatened by management instructions made at L1 and L2. Overall, the similarly categorised governance tensions, including actual and potential ones, perform differently in each case studied. Additionally, the evolving organisational governance of the three cases experience dynamics where their governance tensions drive waves of institutional changes, such as structural or financial reforms.

7.1.2 How are the Governance Conflicts Reflected and Balanced in Evaluation Mechanisms?

The three case studies showed that their existing individual evaluation activities were gradually becoming **integrated towards an overarching evaluation mechanism that is systemically designed and used for organisational governance**. On the one hand, individual evaluation procedures, such as the evaluation of projects, programmes, and institutes, became widely formalised and professionalised at the PRIs, thereby meeting international standards. Meanwhile, the purposes for using evaluation mechanisms to advance organisational governance were confirmed by the analyses of evaluation policy documents and interviews at MPG, HGF, and CAS. The following is a summary of the common purposes of the evaluation mechanisms at the three PRIs:

- ① Assuring scientific excellence by internationally-renowned peer reviews of various research units: projects, programmes, centres, and institutes;
- ② Justifying public funding by publishing summaries of selected evaluation reports addressing the achieved organisational missions and strategic goals;
- ③ Ensuring efficient resource allocation among the research units that compete and collaborate with each other, or among the existing and future research fields;
- ④ Providing opportunities for interaction and negotiation for various stakeholders in evaluation processes (guideline making, on-site inspection and feedback on results) to make evaluation more transparent, trustworthy, and reliable.

According to comparative analyses of the three cases, the above common evaluation purposes were found to actively respond to the corresponding

governance requirements as well as tensions in individual cases. Towards the above purposes, individual evaluation activities at the three PRIs are systemised to supplement each other in terms of levels, subjects, timing, stakeholder participation, and uses. How the three overarching evaluation mechanisms respond to their own governance requirements was presented through the *seven dimensions* in the spider chart (Figure 6.4), such as strategic and financial uses by governance. Overall, the person-centred governance with a high level of scientific freedom and trust at MPG is legitimated by highly international and professional evaluation of individual institutes and the leaders. The evaluation of *Programme-oriented Funding (PoF)* at HGF guarantees the coherence, consistency, and concordance of the collaborated programmes across centres. The *One–Three–Five evaluation* at CAS was designed and is used to integrate fragmented evaluation activities to achieve the aggregated governance of heterogeneous research at its 104 institutes.

Across the three case studies, the systemic nature of the three evaluation mechanisms all experienced institutionalisation processes (Figure 6.5) with their different speeds and scales (Figure 6.6). Such institutionalisation processes, including both evaluation procedures and cultures, were traced from some selected elements (Table 6.2), such as quantity and quality of evaluation guidelines, and having an internal office responsible for organisation and coordination of evaluation activities. It was found that **institutionalisation of the evaluation mechanisms helped to reflect and balance** the governance tensions involving various stakeholders. For MPG, with the least tension observed among the three cases, the strictly implemented evaluation procedures embraces a culture with a high level of acceptability and credibility of evaluation results, because the results come from *the best evaluators* worldwide in particular research fields. For HGF, the three rounds of PoF evaluation, always focusing on the alignment of strategic goals, have increasingly considered the centres' autonomy by coordinating with the midterm evaluation that is empowered to the centres themselves. CAS has experienced a series of reforms to cope with governance tensions, including the increasing use of international and professional evaluations.

Furthermore, patterns of *stakeholders' participation and negotiation* in evaluation activities at the three PRIs were considered by this thesis as the *balancing processes* to mitigate the governance tensions. The patterns observed at the three cases were quite different: the bottom-up style at MPG (Figure 6.7), the

programme-centred style at HGF (Figure 6.8), and the emerging but still mixed style at CAS (Figure 6.9). However, some common principles were found at various phases of evaluation procedures, such as stakeholders' negotiations in evaluation guideline making, on-site inspection, and releasing results.

Consequently, the two ways in which evaluation mechanisms facilitate organisational learning were confirmed at the three PRIs. The first is that the three types of knowledge bases (analytic, synthetic, symbolic) produced from evaluation findings are used for governance (section 6.4.1). The analytic knowledge bases with detailed scientific recommendations are applied exclusively to improve the individual evaluated subjects. The synthetic knowledge bases contribute to synergies across several evaluated subjects for institutions' strategic development. The symbolic knowledge bases are used to verify the credibility of the evaluated subjects. The second way is by the evaluation mechanisms' own learning from their institutionalisation experiences and from stakeholders' interaction and negotiation (section 6.4.2). As a result, the more comprehensive and interactive evaluation mechanisms gradually drive the PRIs' governance to be more responsive to diverse stakeholders—their requirements and tensions (see Figure 6.3).

7.1.3 What are the Balanced Results?

As a result of the above balancing processes, the overarching governance targets for MPG, HGF, and CAS were generalised as **tailored governance arrangements for various research fields, centres, and institutes under the organisational umbrellas**. This overarching target is integrated by this thesis from the ideal balance effects at three organisational levels, taking into account the definition of the balancing role of evaluation mechanisms in Chapter 2 and the identified corresponding tensions in Chapter 5:

- ① The balance between external justification of public funding and internal governance of heterogeneous research at L1;
- ② The balance between organisational strategies alignment and institutes' autonomous development at L2;
- ③ The balance between meeting management instructions and protecting scientists' trust and freedom at L3.

The ideal *balanced results* at each organisational level were interpreted by the author based on the individual and comparative case studies. Most of the ideal governance targets have not been fully realised at the three PRIs; some targets are still at early stages of development. That is why the word *target* is used, instead of *effect* as proposed in Chapter 2, to describe the balanced governance results. For instance, the protection of scientific freedom at L3 was agreed on by the three PRIs to be conditioned by a series of evaluation procedures—rigorous guidelines, strict implementation and the restricted authorities with access to evaluation results. However, the evaluation practices at the three PRIs are much more complex than just following those procedures. So far there is no profession of *moderator* or *mediator* at each of the three PRIs to particularly deal with governance tensions. However, the interviewees were found to have a growing belief in the contribution of evaluation mechanisms to realising the above governance targets.

The ideal effect of governance equilibrium for the three PRIs was interpreted by this thesis not referring to a fixed and satisfactory condition, but to organisational governance developing in a stable manner towards the target of fewer tensions. It should be noted that all these findings were conditioned on the fact that MPG, HGF, and CAS had already achieved successful metrics-based evaluation, according to the *Nature Publishing Index 2013*, as introduced at the beginning of this thesis. Moreover, the costs of designing and using overarching evaluation mechanisms are seriously considered by all the cases. Yet, return on evaluation investment is highly context-specific and path-dependent (Arnold, 2004, p. 14). How their returns match costs in each of the PRIs studied is beyond the scope of this thesis.

7.1.4 What is the Generalisable Principle?

To answer the fourth sub-question, a generalisable principle beyond the three different cases was interpreted as a proper **contextualisation of evaluation design and use**. This principle was derived from analysing both theoretical literature and empirical cases. In the literature related to the concept of *contextualisation*, contextualism describes a collection of views in philosophy that emphasise the context in which an action occurs and can only be understood relative to that context (Price, 2008). Lundvall (1992) proposes that in national innovation systems interactive learning between the innovation units should be contextualised institutionally and culturally. According to Peters and van Nispen (1998), the institutional choice of policy instruments is conditioned by context-specific

appropriateness of the instrument as a result of the context's dependency on the effect of instruments. These discussions can also be applied to PRIs. A lack of consideration of institutional contexts in evaluation studies is recognised by some scholars (Edler, 2012; Hasson, 2006; Højlund, 2014; Raina, 2003). In particular, Whitley and Gläser (2007) emphasise that institutional change and learning related to research evaluation cannot be adequately understood without locating them in specific contexts of research institutions. Therefore, the study of evaluation mechanisms at PRIs defined by this thesis, including purposes, procedures, uses, and impacts, should also consider the national (or EU-like system) and institutional contexts.

The definitions of the two concepts, their characteristics indicators, and related analyses, seriously took the national and institutional contexts into account. To be specific, the *balancing processes and results* of the evaluation mechanisms at MPG, HGF, and CAS depend on the identification of their particular *tensions* (Chapter 5) involving the stakeholders in the national and institutional contexts (Chapter 4). The context analyses provide the basis to understand and explore the research questions and propositions within the PRIs. Indeed, the overarching research question and the key proposition discussed in this thesis are barely defined in any fixed form but only in an interpretable way. The generalisable principle found beyond the three PRIs is the key to understanding and applying the balancing role of evaluation mechanisms for PRIs in general, that is, to contextualise designs and uses of evaluation mechanisms to mitigate context-dependent governance tensions. Therefore, the analytic generalisability, that is, the ability of the findings to explain phenomena in other contexts (Yin, 2009), is enhanced.

How can this generalisable principle be operationalised? In fact, the national and institutional contexts of the three cases analysed were specified by the *eleven dimensions of organisational governance* (Table 4.1), such as research orientations, governance aspects, internal and external stakeholders, and knowledge bases. Meanwhile, the seven dimensions of evaluation mechanisms (such as for strategic and funding uses in Figure 6.4) aggregated characteristics of individual evaluation activities (level, subject, timing, participation, and use). In this sense, the contextualisation principle to explore the linkages between the two core concepts—organisational governance and evaluation mechanisms—can be operationalised as **designing evaluation dimensions aligning with governance dimensions**. For instance, *stakeholder* was found to be a connection between the

three concepts: the *multi-stakeholders* dimension in organisational governance, the *participation* dimension in evaluation activities, and the *stakeholders' requirements* dimension in evaluation mechanisms. Therefore, the design of stakeholders' participation in evaluation activities and of how their requirements influence overarching evaluation mechanisms should align the interests and power of particular stakeholders involved in organisational governance. Afterwards, the *strategic and financial uses of evaluation results*, as dimensions of both evaluation activities and mechanisms, explicitly align the governance dimensions of *organisational strategy* and *funding framework*. This alignment was interpreted as contextualisation processes of evaluation design and use.

Although some dimensions of organisational governance were found to be aligned with some dimensions of evaluation mechanisms, neither the concept of typology nor typological theory (Højlund, 2014) was used to study their linkages because there are only three cases in the thesis. Each case performs individually under the contextualisation principle, which does not make much sense for a typology (George & Bennett, 2005). However, more alignment patterns between organisational governance and evaluation mechanisms are expected to be found beyond the three cases, which leaves open-ended discussions for PRIs in general.

The results of testing the four propositions of the enriched conceptual framework (Figure 4.3) in the case studies have been presented in sections 7.1.1 to 7.1.4. Based on that, the final conceptual framework as the analytical instrument of the whole thesis is discussed in the next section.

7.1.5 The Completed Conceptual Framework

The final conceptual framework, as shown in Figure 7.1, aims to illustrate the completed rationale of the balancing role of evaluation mechanisms in organisational governance of PRIs. It is based on the theoretical formation of the basic framework (Figure 2.1) and then enriched by the case studies of MPG, HGF, and CAS. This analytical tool has been proved empirically to be useful for highlighting the **co-evolution of organisational governance and evaluation mechanisms at PRIs**. Supported by the empirical findings, how PRIs are positioned in national contexts is important since national research systems laid the groundwork for the public justification that PRIs have to respond to. Multiple stakeholders in national systems (such as funding bodies and scientific

communities) often have conflicting expectations concerning PRIs' relative roles and performance. In this context, evaluation appears as an often-chosen approach for PRIs to justify their public funding and scientific excellence. To zoom into institutional contexts, internal stakeholders at the three organisational levels (decision-making and supervisory body at L1, research institutes at L2, and individual researchers at L3) at the central position of the framework, decide and participate in both governance and evaluation activities.

The roles of stakeholders across the three levels (L1, L2, L3) were confirmed as the core factor in the **two bridges linking organisational governance and evaluation mechanisms**, as the green arrows show in Figure 7.1. Stakeholders' tensions about governance aspects (strategy, funding, operation) justify the purposes of evaluation mechanisms, as represented by the upper bridge in the figure. Evaluation mechanisms were found to actively respond to governance requirements involving multiple stakeholders and to provide evidence to mitigate their tensions, which was supported by empirical findings in section 7.1.2. To do so, evaluation mechanisms should experience institutionalisation processes to become more professional and systemic in terms of procedures, and more responsive and interactive in terms of stakeholders' participation and negotiation. Such institutionalisation processes, as well as learning processes of evaluation mechanisms, go through the diffusion of both evaluation procedures and cultures on people's acceptance of the credibility of evaluation results. Afterwards, evaluation results generate three types of knowledge bases (analytical, synthetic, symbolic) for governance uses and impacts, which facilitates organisational learning as shown by the lower bridge in the figure. In principle and practice, such co-evolution of organisational governance and evaluation mechanisms is also a dynamic learning process. The balanced governance results after the evaluations' functions again impact the next round of evaluation activities and uses. The co-evolution between organisational governance and evaluation mechanisms finally becomes evident.

Figure 7.1 The completed conceptual framework on the role of evaluation mechanisms in organisational governance

Source: the author, 2015

An abductive process has been taken for completing this conceptual framework as an analytical tool for PRIs in general, which is a key result of this thesis. The basic framework only had the primary heuristic of the relationship between organisational governance and evaluation mechanisms based on limited literature. Specific *bridges* between these two concepts were considered as two empty boxes. Then the elements on governance tensions—stakeholders’ interests in governance aspects (strategy, funding, operation) and power at organisational levels—were added to the first box after analysing them empirically. Next, the common principle on how evaluation mechanisms respond to and mitigate governance tensions was found, based on the three different cases, which helped to fill the second empty box.

To apply this conceptual framework for PRIs in general, the entry points can be the characteristics of individual elements of organisational governance and evaluation mechanisms, as shown by the blue boxes in Figure 7.1. To explore the relationship among these elements, interactions of various stakeholders should be considered with special attentions given their interests in governance aspects and power at organisational levels, as shown by the orange arrows in Figure 7.1. Next, the linkages between governance and evaluation, as shown by the two green arrows, can be explored concerning the questions—what conflicts need to be balanced and how evaluation can be used as governance instruments to mitigate those tensions. In this exploration process, dynamics of organisational governance and evaluation mechanisms need to be reviewed. This conceptual framework can be used to open more black boxes about evaluation design and use as governance instruments than those of the three PRIs studied. In addition to this contribution, other theoretical contributions of this thesis will be presented in the next section.

7.2 Theoretical Contributions

Overall, the understudied literature on the **relationship between organisational governance and evaluation is enriched** by this thesis, through an overarching conceptual framework (Figure 7.1) and new elements or perspectives added into the two concepts. For instance, new perspectives of institutes’ autonomy at L2 and

of scientists' trust and freedom at L3 can be paid further attention by the current organisational studies (organisational governance and learning). Organisational theories of governance and learning in general (Hermanson & Rittenbery, 2007; March, 1991) helped to interpret and analyse their characteristics at PRIs that were rarely discussed. In turn, the theories are widened by findings at PRIs. Furthermore, the findings of governance conflicts contributed to stakeholder theory (Freeman, 1984) through confirming the three-level organisational structure (Hage et al., 2007) and stakeholders' diverging interests and power (Bryson et al., 2011) across the three levels. The tensions identified at PRIs strengthen the application of stakeholder theory in studies of research institutions.

Moreover, the concept of *evaluation mechanisms* is defined and developed by this thesis towards its systemic nature involving individual evaluation activities that are often discussed but disconnected in evaluation studies. Individual theories on evaluation, their subjects and objectives as discussed in section 2.3.1, can be systemised by *evaluation mechanisms* to realise their balancing roles vis-à-vis complex governance and stakeholders' tensions. This notion and its design and use aligning organisational governance—the contextualisation principle in section 7.1.4—are innovative and can become an item combining studies of evaluation and organisation. For instance, beyond the normal dimensions of evaluation activities (such as level and subject), elements of an evaluation mechanism, such as stakeholders' requirements in evaluation guidelines and stakeholders' authorities on access to different versions of evaluation results, reflect evaluation mechanisms' nature as a governance instrument.

This thesis contributes elements to trace *institutionalisation processes of evaluation mechanisms* within PRIs (Table 6.2), such as quantity and quality of evaluation guidelines, and diffusion of evaluation procedures and culture. As supported by section 2.2.2, the current evaluation literature discusses institutionalisation of evaluation mainly at national levels, such as university systems; PRIs are rarely involved. However, evaluation institutionalisation at organisational levels justify the balancing as well as learning processes that individual evaluation activities in one PRI can develop to complement each other. These processes also prove that, in addition to reliable evidence and legitimation, evaluations increasingly facilitate stakeholders' negotiation to mitigate governance tensions. *Negotiation*, as a key concept of the Fourth Generation Evaluation (Guba & Lincle, 1989) and the Empowerment Evaluation (Fetterman et al., 1996), can be

applied beyond individual evaluation subjects (e.g. programmes) but also in relation to integrated subjects like research institutions.

Additionally, the *principles on why and how the balancing role of evaluation mechanisms* taking place at PRIs contribute to system-level studies of governance instruments of research and innovation. For instance, national innovation systems, which are pictured as a process of communication and interactive learning among innovators (Asheim & Gerlter, 2005; Johnson, 1998; Lundvall, 2007), can integrate the balancing and learning processes in individual innovators. The relative roles and tensions among various stakeholders within one system can be mapped and considered in a systemic way via some kind of governance instrument, like evaluation mechanisms. Moreover, the challenging balance between exploration of new possibilities and exploitation of old certainties (March, 1991) can be pursued by developing and using knowledge bases that are needed. Evaluation mechanisms are confirmed to generate such knowledge bases (analytic, synthetic, symbolic) used by PRIs' governance. These three types of knowledge bases, originally from studies of regional innovation (Asheim et al., 2011), can be used as a new element when looking for or testing governance instruments of research and innovation at system levels.

Furthermore, this thesis tries to distinguish **the intended and unintended effects of the balancing roles of evaluation mechanisms**. However, this is still an open-ended question and provides another tension to be balanced. This thesis studies how overarching evaluation mechanisms are aligned with multiple stakeholders to mitigate their conflicts. The formal and intended evaluation procedures analysed by this study are inevitably mixed with spontaneous or unintended activities of stakeholders. For instance, stakeholders' interaction taking place in one evaluation activity results in changes in their behaviours in other activities, which are not traced by this study. The intentional or unintentional behaviours of stakeholders also depend on the availability and transparency of the evaluation information that is communicated. Therefore, more diverse and informal inputs of stakeholders, such as networking, can be considered when identifying stakeholders' tensions as well as ways to mitigate them. Knowledge can be drawn from the results of this thesis that are undervalued by literature.

7.3 Recommendations

The general and complex questions asked at the beginning of this thesis—how can research be evaluated, and can there be standardised research evaluation within PRIs—cannot be fully answered by this thesis. However, this thesis contributes the experiences of leading PRIs worldwide (MPG, HGF, CAS) and the following recommendations about what can be learnt from them.

7.3.1 Implications for PRI Practices

The results of this thesis provide **explicit recommendations** for PRIs beyond the three presented cases. An analytical instrument (Figure 7.1) and details of how it is used at MPG, HGF, and CAS are offered for PRIs to design and use their own evaluation mechanisms for organisational governance. The ways to analyse governance characteristics and identify governance tensions can be applied for PRIs to find out their own problems. For instance, the three-by-three matrix considering governance aspects (strategy, funding, operation) and organisational levels (L1, L2, L3) can be applied. The governance tensions found at MPG, HGF, and CAS are summarised as a simplification of a much more nuanced picture of their practices. Those examples can help PRIs to prevent predicted tensions, or to learn how to solve them by knowing what they are or might be. If evaluations are not capable of responding to certain tensions, they can also have minor functions, for example, to provide a space for stakeholders' negotiation or a third-party perspective for governance to find another solution.

For a large-scale PRI with rich resources like the cases studied, it is suggested that **governance requirements are aligned in the design and institutionalisation of an overarching evaluation mechanism**. In principle, individual evaluation activities at one PRI should be designed to supplement each other in terms of dimensions, for example, timing (*ex-ante* and *ex-post*), scale (big and small), and frequency (*ad hoc* and regular). At best, comprehensive evaluation mechanisms cover a wide or even the full range of the constituted elements of those dimensions for the systemic nature of evaluation mechanisms. Some dimensions related to governance requirements are not easily changed in a short time but should be noted at the very beginning when designing evaluation mechanisms and be kept institutionalised, such as strategic and financial uses of evaluation results. Some increasing public concerns for PRIs, such as societal participation and impact of

research, can become new dimensions in existing evaluation mechanisms. It is important to inspect new, and update existing, dimensions regularly. An overarching evaluation mechanism must experience institutionalisation dynamics influenced by multiple elements, such as diffusion of evaluation procedures and culture on people's acceptance of evaluation results. A special and independent department in the headquarters should be responsible for the coordination of various evaluation activities for their systemic uses. More freedom for and trust of researchers is suggested rather than relying on vigorous procedures and criteria.

The contextualisation of evaluation design and use for aligning governance (section 7.1.4) can take various forms, like the large-scale structural reforms at CAS or changes of funding programmes at HGF. At MPG, in contrast, there are limited governance uses of evaluation results because of its governance tensions which are found to be the least among the three cases. Such different situations support the common principle of contextualisation of evaluation design and use considering their different tensions. In this sense, this common principle found from the three PRIs can be applied in various forms for PRIs in general. However, for PRIs that do not have as large-scale or successful metrics-based research evaluation as the three cases, their single or limited numbers of evaluation activities probably develop quite individually. The idea of an overarching and learning evaluation mechanism as a governance instrument is not recommended as a short-term target for PRIs at early stages of institutional development.

Furthermore, it is recommended for PRIs including CAS to **increase the participation and interactions of stakeholders** in evaluation procedures. The quality of stakeholders' interactions can be improved through more inclusion of key actors and their deliberation in communication (Kuhlmann et al., 2015). When doing so, supportive conditions should also be matched. For instance, evaluation data collection and analytical processes can be more transparent, at least within PRIs, to facilitate the learning between the evaluated subjects and stakeholders. Stakeholders' engagement and interaction in evaluation activities can be designed based on their diverging concerns by answering the question of what information can be shared by which stakeholders. Since there is always limited transparency of evaluation results for PRIs, both MPG and HGF prepare various versions of evaluation results with different degrees of confidentiality for different stakeholders whose access to corresponding information is dependent on their power. For instance, general summaries are available for a wide range of

stakeholders; detailed results have limited access. This can be understood as a balance between transparency of evaluation guidelines, to ensure the objectivity of evaluation implementation, and confidentiality of evaluation results, to protect scientists' trust. This way works well in MPG and HGF, which already have categories of evaluation results and a mature culture of communication of different categories. For those PRIs that don't, it is an important recommendation. For instance, more transparent information on evaluation guidelines and results at CAS are expected by the author and some internal staff.

PRIs, including the three cases, are suggested to categorise and aggregate *knowledge bases*, generated by evaluation findings, to govern heterogeneous R&D across various fields and units under one organisational umbrella. The premise is that competitive advantages of heterogeneous fields and units should be protected, such as developmental autonomy of institutes at L2 and creativity and trust of scientists at L3. Based on that, scientific knowledge generated from individual evaluation findings can be aggregated and systemically used for high level decision-makers, such as the added value of synergetic knowledge of multiple evaluated subjects. Extended evaluation at MPG, Programme-oriented Funding (PoF) evaluation at HGF, and One-Three-Five evaluation at CAS all have the strong intention to create more collaboration and synergies among their core research units. For the three cases, such ideal synergies still need to be realised in governance practices beyond the evaluations' intentions. For other PRIs, which may not have particular synergy-aiming evaluations, actions can be started from analysing individual evaluation findings and ensuring that complement each other. Therefore, PRIs can learn to be more adaptable to the changing environment and take systemic and synthetic decisions to achieve organisational missions.

7.3.2 Reflections on Public Policy of Research Evaluation

Some reflections about public policy of research evaluation, beyond the research question of this thesis, are discussed in this section. These reflections are derived from the three disconnected parts: a) literature discussion on system-level tensions (section 2.2.1); b) national R&D landscapes of Germany and China (section 4.1); c) governance tensions found at L1 of MPG, HGF, and CAS (section 5.1). Based on the three sections, public policy-making on research evaluation is suggested to seriously consider the two indexes (axes) in Figure 7.2. This suggestion aims at more integrated and better-categorised governance of public R&D at national (or

EU-like system) levels. So, where are Germany and China positioned in Figure 7.2, and why? Should they change their current positions?

Figure 7.2 Reflection on public policy of research evaluation

Source: the author, 2015³³

The two indexes in Figure 7.2 are not totally independent of each other because they both depend on scientific and economic levels of a nation or system. The horizontal axis refers to the intensity of R&D expenditures for a nation or system. According to section 4.1, both Germany and China have high levels of R&D investment. The vertical axis indicates the awareness and capacity of public policies to categorise the multiple research orientations (fundamental, strategic, applied) of the R&D institutions in a R&D landscape. Governance of the comprehensive R&D landscape should have a clear understanding of what is happening on the ground for each research orientation (Rip, 2000). Inspired by the work of Crow and Bozeman (1998), national policy-making according to categorised research orientations and the involved institutions can be effective only if the existing R&D landscape comprehensively covers all research orientations and develops to be diversely productive. Germany is fully qualified for such conditions since the

³³ Drawing this table was inspired by the book of Hoppe (2009).

interviewees at MPG and HGF showed their positive attitudes to the categorised evaluation policies made by the Federal Ministry of Education and Research (BMBF), such as the targets of international excellence for fundamental research and national usefulness and relevance for strategic research. Therefore, Germany is positioned by this thesis to have ideal *integrated* public policies of research evaluation, as shown by the upper right cell in Figure 7.2.

Reading the four categories formed in Figure 7.2, public policies of research evaluation are suggested to perform high on both of the two indexes to achieve the ideal *integrated* policies. When performing low on both of the dimensions, the quantity and quality of public policies on research evaluation are assumed *small* and *weak*. Even with higher investment, public policies of research evaluation would only be expected to have *low efficiency*; the low awareness and capacity of policies to categorise various research orientations cannot guide resources to be allocated efficiently to R&D fields and institutions. It is likely that each orientation develops autonomously and isolated from others, leading to a complex and inefficient operation between the orientations. The typical lesson is from the experiences of CAS. Even though it has high R&D investment, its research orientations develop fragmentally and need interdisciplinary collaboration. That is why China's research policies, as well as CAS itself, have undertaken many reforms in recent years aiming at clearer classification of the mixed R&D landscape. Therefore, China is now positioned in the *low efficiency* category in Figure 7.2 but making efforts to move to the category of *integrated*.

When the categorising level is high but with limited R&D resources, low effects are expected because each research orientation cannot develop fully towards its potential without enough funding. The ideal *integrated* policies are supposed to function in a clearly categorised and highly invested R&D landscape embedding various R&D orientations and the involved institutions. This blueprint may only be applicable for strong R&D countries with comprehensive research orientations. To achieve the *integrated* research policies, multiple research evaluation tools, for example, metrics-based evaluation and qualitative expert assessment should complement and coordinate with each other towards the target of an overarching and systemic evaluation mechanism. Furthermore, public policies on research evaluation should be a part of evolving systems considering emerging complex challenges: new societal goals, new stakeholders, and system-level tensions

between competition and collaboration across research orientations, fields, and institutions.

All the results and implications of this thesis have been presented. The next final two sections will discuss the limitations of this research and ideas for future research.

7.4 Limitations of This Research

Since social scientists are embedded in the intersubjective social processes of the worlds they study (Yanow, 1998), the ways to interpret and analyse qualitative data in this study inevitably have limitations. Objectivity is the first concern. The selection of the data resources at each level of each PRI analysed has subjective considerations: who to talk to, which data to choose, and how to interpret it different kinds of arguments. There was very limited access to the high-level decision-makers at L1, like the president, vice president, or even department heads, of the three cases. Some attempts to approach the key stakeholders were refused or unanswered. The selection of the interviewees at L2 and L3 was not perfectly controllable either. The numbers of different types of interviewees accessed were not balanced, for example, MPG did not have as many interviewees at L2 as did HGF and CAS (Table 3.9). In addition, some potentially useful data, such as evaluation results, cannot be collected because of confidentiality reasons.

Furthermore, it may not be possible to recognise the possible failures or side effects of the evaluation mechanisms. Understandably, individual interviewees interpreted the research question and the key concepts of this study differently according to their hierarchical levels and roles. They talked about their own judgement, perceptions, and impressions based on formal policies on governance and evaluation, and informal communication with other stakeholders. There is a risk that some valuable evidence beyond the defined dimensions (of governance and evaluation) might have been missed. For instance, the identification of governance conflicts, rather than being linear and sequential, was an iterative and interactive process. It is likely that only small parts of the black boxes of the governance of the three were opened.

Additionally, the diversity of cultural understanding of the key concepts—*tension*, *trust*, and *balance*—was not systemically studied in this thesis. Such diversity was identified at the beginning of this research and afterwards when translating and

interpreting qualitative data. The recognition of these cultural differences was helpful to generalise conceptually and empirically the categories of governance conflicts and the roles of evaluation mechanisms. Also, the cultural divergence indicates that the relationship between *tension*, *trust*, and *balance* is neither uniform, nor linear, nor deterministic. The differences of national R&D landscapes in Germany and China support the need for policy studies to be rethought, using the question, how can credibility be established across international public arenas (Beck, Bonss, & Lau, 2003)? Undertaking this study across the two subject countries (Germany and China) in a third country (the Netherlands) was helpful to identifying the patterns to achieve independence and objectivity in international spheres. Additionally, many third-party researchers consulted verified these differences.

This study takes place at a time when CAS is changing rapidly. Some important reforms, such as the *One–Three–Five evaluation*, were already involved in the discussion on the evolving governance of CAS. However, not all the emerging or ongoing changes of internal organisations and initiatives for governing complex R&D are involved in this thesis. For instance, the CAS S&T Strategy Consultancy Research Centre was established in June 2015. It is closely relevant to CAS’s strategy-making but it is impossible for the author to follow all these changes given the limited time and resources of this research. Some important relevance of those changes to this thesis might have been missed and can be considered in future research.

7.5 Future Research

Future research aims to overcome the above limitations of this study. Some propositions of this thesis will be tested in a wider range of international cases. Dimensions of organisational governance and evaluation mechanisms can become more diverse and advanced with a wider selection of PRIs. In particular, stakeholders’ responses about those dimensions will be investigated, which may widen and alter the current definitions and descriptions. Different cultural understanding of organisational governance and evaluation mechanisms and their impacts will be explored in other countries beyond Germany and China. This calls for researchers to contextualise and operationalise some abstract concepts, like scientific freedom, trust, and responsibility of scientists. In that sense, the common principle—contextualisation of evaluation design and use—will be applied more

widely. There can be more ideal types of contextualisation of evaluation design and use if more PRIs are included in future research.

Additionally, some potential effects of evaluation will be tested when more evidence is collected over a longer period. For instance, some predicted effects and impacts of the ongoing *One–Three–Five evaluation* in CAS will be measured in another two years and further in the future. A new diagram illustrating the research system of CAS involving the newly developing four R&D clusters, similar to Figure 4.1 involving MPG and HGF, is expected to be drawn several years from now. In addition, it would be worthwhile to explore the future attempts of the three PRIs to change their existing governance characteristics and balances in favour of new or alternative ways of evaluation.

Moreover, the current discussions on co-evolution between organisational governance and evaluation mechanisms could be applied for the general governance of research beyond organisational levels. An extended topic can be evaluation as a means of responsible governance of emerging R&D systems. This topic links with the governance tensions found in this thesis at L1 about public justification of research investment, with reflections on public policies of research evaluation, and with the increasing need for scientific trust and freedom for scientists at L3. For instance, various stakeholders' tensions should be anticipated at the design phase of evaluating emerging R&D systems in a responsible way. The balancing roles of evaluation mechanisms, particularly their institutionalisation, could become the basis of evaluating and governing emerging R&D systems. The governance of responsive research and innovation (RRI) is becoming a central feature of research and innovation policy in Europe and it is still evolving (Walhout & Kuhlmann, 2013), addressing grand societal challenges towards a *responsibilisation* (Dorbeck–Jung & Shelley–Egan, 2013). Governance of RRI creates spaces of openness, probing, and learning instead of limiting options for actors, institutions, and processes (Kuhlmann et al., 2015). This thesis makes similar suggestions to PRIs on how to govern and evaluate research and innovation responsibly, including to increase stakeholders' participation and the quality of their dialogue, and how to encourage their open-minded responsiveness and forward-looking attitudes.

Appendices

A MPG

The following supplementary materials of Max Planck Society (MPG) were collected and interpreted explicitly from its official website (<http://www.mpg.de/en>) and two published policy documents—2013 annual report (MPG, 2013) and evaluation documents (Schruff, 2012). This section also includes primary data (interview findings) that were used in the thesis.

A1 Historical dynamics

MPG was founded in 1948 as the successor institution to the Kaiser Wilhelm Society established in 1911 and named after Physics Nobel laureate Max Planck. When established, it comprised 25 institutes and research centres. In 1949, even before the Federal Republic of Germany was established, the federal states guaranteed that the Society received sustained institutional financing. Since then, its collaboration with institutions outside of Germany has gradually been cemented. New research centres were established for biology and biochemistry during the 1960s. The spectrum of research in the fields of physics and chemistry expanded to include astronomy and solid state physics. Research in the humanities and social sciences were started on socio-political questions. By 1966, the number of research institutes had grown to 52. In the seventies and eighties, MPG concentrated on interdisciplinary cutting-edge research in particularly fields and by specific programmes. During German reunification, 18 new institutes were set up in eastern Germany before 1998.

A2 Organisational structure

The organisational structure of MPG is shown by Figure A2. The Executive Committee advises the president and draws up the overall budget and prepares the annual accounts. The Senate is composed of the president and a minimum of 12 and a maximum of 32 elected Senators. The Senate elects and appoints the president (for a six-year term), members of the Executive Committee, the Secretary General, Scientific Members including institute directors, and Supporting Members. The Senate also deliberates on the establishment or closure of Max Planck Institutes (MPIs), the MPG's involvement in science and innovation systems nationally and internationally, the strategies and entire budget. The Scientific Members of MPG discuss matters of common interests and make recommendations to the disciplinary Sections. *Ex officio* Senators include the president, the chairperson of the Scientific Council, the chairperson of each of the three Sections, the Secretary General, three scientific staff members chosen by each Section, the chairperson of the general works council, as well as five ministers or undersecretaries representing the federal and the Länder governments. The Honorary Members and the Honorary Senators are also members of the Senate and have an advisory capacity.

Figure A2 Organisational structure of MPG

Source: MPG website, 2013

The Administrative headquarters of MPG in Munich is headed by several Secretary Generals. It runs the daily business and administrative tasks, advises and supports MPIs and research facilities in preparing and implementing decisions. The Unit of Institute Development and Evaluation is responsible for organising institute evaluation procedures. The Scientific Members are appointed by the Senate, including the Scientific Members (usually the directors) and the External Members of MPIs. They participate in the meetings of the Scientific Council in an advisory capacity once or twice a year. They discuss matters of common interests of and make recommendations to the Sections (biology and medicine, chemistry, physics & technology, and human sciences). The Board of Trustees deliberates on science policy as well as on economic and organisational questions, additionally facilitates contacts to those circles interested in MPG.

A3 Strategy, funding and operation

MPG regularly takes a position, often together with other scientific organisations, on research policy issues like research promotion in Germany and Europe, and developments on specific topics. The record of successes of MPG is reflected by the large number of famous, award-winning scientists, 18 Nobel laureates so far, over 15,000 publications each year in internationally renowned scientific journals, and third place in the *Nature Publishing Index* (2013).

MPG creates opportunities for young researchers' international scientific careers. Now there are more than 9,000 junior researchers working at MPIs. The International Max Planck Research Schools (IMPRS) since 1998 in collaboration with universities have motivated talented PhD students to come to Germany for postdoctoral studies. Each year MPG awards the Otto Hahn Medal for outstanding PhD theses by providing financial assistance for a postdoc position abroad. The independent Junior Research Group at MPIs let young researchers lay the cornerstone for their future scientific career with five years to pursue their own goals on a limited but secure budget. These highly coveted positions are announced internationally and awarded in a competitive process. Junior scientists are offered excellent research opportunities. MPG is proud of educating the best future scientists who work inside or outside its own institutes.

MPIs are involved in more than 2,000 cooperation projects with nearly 6000 partners in over 100 countries. More than 6,000 foreign visiting and junior researchers work at the institutes each year. A third of the directors, half of the PhD students, 80% of post-docs are not German. There are now more than 40 partner groups working in Asia, Eastern Europe, and South America, like the Partner Institute in Shanghai with CAS, the Bio imaging Institute in Florida; and there are several plans for the future.

In the past 20 years, MPG has advised 86 spin-offs and generated revenues of around €200 million for inventors and MPIs. To accelerate the transfer of promising findings in basic research to potential applications, joint efforts with other institutions are made. MPG has recently begun expanding its cooperation projects with FhG in certain fields to economic success, such as computer science, materials science, nanotechnology, biotechnology, and regenerative energies. The two institutions have been engaged in talks since 2004 to identify and support collaboration at the interface of basic research and applied research.

Box A3 Supplementary interviews at MPG institute A

Interview with a senior scientist:

Our educational and public outreach activities go beyond the ordinary, looking back on a long and successful tradition, from media outreach to successful public lectures and to large-scale events such as Open Days hosting over 10,000 visitors. This tradition has continued in a key target group of young students audiences. An example is Wissenschaft in die Schulen [Science into our schools], a highly successful joint venture between our institute and Germany's most popular magazine in our field.

Interview with an evaluation organiser:

The external dialogues with other institutions in the same research field are the responsibility of researchers who come across peers all over the world and disseminate recipes [knowledge], in terms of joint professorships with universities, conferences, workshops, and also forums for early-career researchers. The mobility of excellent researchers, especially those going somewhere else after studying in MPis, is paid great attention by us to enhance research connectively.

A4 Evaluation practices

In *ex-ante* evaluations, four years before a Scientific Member being due to retire, the responsible institute is asked to propose a concept for the future research and the potential successors. The relevant Sections of the Scientific Council would set up an appointment committee composed of Scientific Members and external experts. The committee comment and recommend the scientific concept and the long-term prospects, the possible candidates and the required resources. After identifying a potential candidate, up to 15 written references are obtained from other internationally renowned scientists. Then the proposal is discussed in depth by all of the relevant Scientific Members of the Section. Once the Section has come to a positive conclusion and submitted its recommendation to the president, negotiations are open with the candidate. The following endorsement is by the Senate and the appointment is made by the president.

In preparation for *ex-post* institute evaluations, the institute first submits a status report describing the current status of research, new initiatives, collaborations and publications with international comparison, personnel deployments, budget and acquisition of third-party funding. Particularly, it also documents the subsequent positions held by staff after leaving the MPIs, thus illustrating the career development opportunities for junior scientists. The material prepared by the institute, ranging from 100 to 600 pages depending on the institute size, is made available to SAB prior to the Board's on-site inspection. The quantitative data in the status report are basis for the relevant criteria. All aspects would be discussed in depth during the on-site evaluation of two to three days. Afterwards, SAB members give a brief oral report at the end of the inspection to the directors. Usually the president and the vice president of the section of the institute are also present during the SAB meeting to get direct feedback. Then all SAB members contribute to the detailed evaluation report, and the chairperson, appointed by the president, is responsible for finally delivering the report directly to the president. The president forwards the report with his comments to the institutes with potential further actions and allows institutes to comment on the report and/or defend themselves. It might be the case that directors would like to make institutional reports public within institute houses. In principle this report is only to SAB members and high-level heads.

The Senate decides the *SAB rules* after dialogues of diverse Scientific Members. Institutes directors have autonomy in influencing SAB regulations. In 2015, MPG has 810 SAB members of which 85% come from organisations outside of Germany. Less than 1% of all SAB members are also members of other MPIs. More than 300 experts participate in about 40 SAB inspection visits each year. Given the major contribution of SAB as both critics and advisers to MPIs, it is important to keep evaluators' independence to guarantee effective quality assurance of research activities. When selecting SAB members, conflicts of interest such as cooperation, joint publications and student-teacher relationships are considered and, where possible, excluded. If a conflict of interest cannot be avoided after carefully weighing the above considerations, the connection is made clear for all participants at the start of SAB meeting. Statistically more than 97 percent of SAB members are not affiliated with MPG and more than 75 percent come from international organisations. SAB members are appointed for maximum six years and new appointments are to replace part of the membership every three years. As a result of this rotation, generally every SAB meeting is attended by both old and new Board members. SAB rules keep being revised continually, for example, once in 2014 and once in 2015. The Chair of the SAB is now appointed by the MPG president in advance and no longer elected. Rules were added for the Support of Junior Scientists.

In extended evaluations, the Research Field Committee is composed of chairs of the respective SAB members, rapporteurs, president, vice president and Section chair. The criteria applied are scientific achievement, the efficient deployment of resources and institutes' medium-term future prospects, closely around governance aspects that are less scientific and more comparable at cluster level, such as collaboration and training. The Committee checks whether the MPIs take advantage of contexts well in their own environments and give general suggestions about direction of next-round development.

Box A4 An example of a partner institute experiencing different governance

Partner Institute for Computational Biology (PICB) was established in 2005 with 2/3 institutional funding from CAS and 1/3 from MPG, and has 18 Principle Investigators (PIs), 5 research directions, 104 staff, and 150 students. PICB mainly conducts fundamental research, curiosity-driven and person-centred, without strict requirements on research topics. Meanwhile, it follows administrative policies of CAS—staff restriction and recruitment, contract duration, quota restrictions when applying for CAS projects, and financial regulation of the higher-layer decision-makers. It was initially set as a window of mutual learning CAS and MPG about experiences of S&T management and education. In addition to the finance partnership, PICB's requirements for SAB members are not as strict as other MPIs in Germany. One SAB member of PICB is a director of another MPI, which is unique in MPG.

Regarding research evaluation, PICB experiences different requirements from MPG and CAS. Evaluative evidence is slightly different because of the different statistical requirements. A greater difference is identified between flexibility of selection and adjustment of research topics in MPG and research stability valued by CAS. The Managing Director is responsible for organising evaluation and feedback exchange between both institutions. So far, it has done 4 MPG institute evaluations, involving a total of 13 SAB members, zero extended evaluation or bibliometrics. The final result of MPG institute evaluation is only accessible for the PIs. Those results impact internal resource allocation accordingly, usually in terms of running fees, space distribution, and quota of training PhDs. However, PICB did not participate in any institute evaluation of CAS before 2013. The interviewees at PICB were eager to know its position, strength and weakness nationally and internationally. It started to take the One–Three–Five evaluation of CAS in only two research directions, resulting in adjustments in a small and flexible research group. No big change of PICB from One–Three–Five evaluation of CAS was observed.

B HGF

The following supplementary materials of Helmholtz Association (HGF) were collected and interpreted explicitly from its official website (<http://www.helmholtz.de/en/>) and three published policy documents: 2013 annual report (HGF, 2014), HGF Strategy 2009 (Gazlig, 2009) and Helmholtz Roadmap for Research Infrastructures (Gazlig, 2012). Some primary data is also included, especially the unpublished evaluation procedures from interviews.

B1 Historical dynamics

Big science of Germany has experienced over 60 years' dynamics. Due to the restriction by the US and Russia from 1950s to 1980s, the first institutions were founded in 1950s mostly from the nuclear physics. Since then some individual national laboratories started to develop in isolation geographically before the reunited federal states of Germany in 1990. After a period of re-orientation and the foundation of new establishments in many fields, the federal government's funding programmes with a focus on substantive aspects of big science were put on a sound institutional basis.

The reorganisation of the national research centres under the umbrella of HGF took place in 1994 to commemorate the one-hundred year birthday of the great German scientist Hermann von Helmholtz. The initiative was to combine the earlier sixteen individual laboratories covering a broad spectrum together under one roof. Those laboratories were re-organised into member research centres of HGF and began to apply for funding officially and directly from the new federal government. Since then the BMBF has been responsible for organising and evaluating their funding application and scientific performance. At the beginning, those works were not organised well because of the HGF's heterogeneous structure and scientific work demanded professional and high technological capabilities. Then the systemic evaluation by WR required HGF to adopt the fundamental principle of big science to identify systemic deficits produce applicable recommendations to the specific circumstances of individual centres. The first phase of the new HGF brand was from 1994 to 2001 when the funding powers were totally in the hands of the government. The second phase started from 2001 with the emerging Programme-oriented funding (PoF). PoF has developed for three rounds by 2014 and responds actively to Germany's policies.

Box B1 An example of HGF responding to national policies

According to the German EFI report 2013, HGF responds positively to the Energy Transition (Energiewende) policy. After the disaster at the nuclear power reactors in Fukushima in 2011, the German parliament adopted a legislative package on 30 June 2011 on the decommissioning of all German nuclear power reactors by 2022. This Energy Transition involves intensive political and social debates. The evaluation in 2012 by EFI on the responses of PRIs to the new energy legislation showed that a comprehensive coordination of all PRIs exists only in rudimentary form. HGF was the most active PRI, focusing on renewable energy, efficient energy conversion, and nuclear fusion, especially in terms of the programme of Technology, Innovation and Society. Several new strategic Helmholtz Energy Activities have been launched and will take effect in the near future. In the funding period 2010–2014, 42% (€562 million) of the HGF's basic funding resources allocated to energy-related research was granted for activities in the field of nuclear fusion. On the contrary, the necessity of other PRIs to deal with nuclear energy and fossil energy conversion was only to a limited extent. MPG's long-term objectives in sustainable energy supply were not affected by Germany's accelerated phase-out of nuclear energy. Its research activity on future energy supply was focused on nuclear fusion and that was also in close cooperation with HGF.

B2 Organisational structure

The Senate elects president and vice presidents, and deliberates on all major decisions such as recommendations to the financial sponsors on thematic and funding priorities. A Senate Commission is constitute by: one expert for each field, permanent representatives of ministries, members of the Senate representing respective fields and the chairpersons of programme review panels. The central actors are the internal Assembly of Members made up of the scientific and administrative directors of all centres and the external Senate. The Assembly of Members discuss developments and decisions at regular meetings. The administrative Head Office in Berlin represents the Headquarters in this thesis. The Committee of Financing Partners has representatives of the federal and state governments. The steering committees of the fields are formed by the programme speakers and the scientific directors of the participating centres and are responsible for research coordination within the respective fields. All the research activities are organised in long-term programmes in the six particular fields.

Figure B2 Organisational structure of HGF

Source: HGF website, 2015

B3 Strategy, funding and operation

The HGF centres participate in the EU framework programmes, which strengthen their comprehensive capacities of managing HGF programmes. The centres also cooperate with national and international universities and industries. The second round of PoF (2009 to 2013) distributes €9.405 billion to the six research fields. Table B3 demonstrates the top-down approach to develop a programme guided by the Strategic Guidelines and Position Paper.

Programme development	Contents
Strategic Guidelines 5–10 pages	Key scientific areas Structure of the research field General review objective and criteria Five-Year budgetary framework
Position Papers of individual research fields 10–20 pages	Major scientific challenges and overarching goals Contribution to achieve these goals Cross programme research activities Common networking strategies Education, talent management, equal opportunity
Programme proposal 100–140 pages	Centres' participation Cross-centre strategy meetings Joint proposal drawn up by scientists

Table B3 The top-down approach of a programme development

Source: HGF headquarters interview, 2013

In order to track scientific progress of programmes, a largely standardised controlling process is made by annual process reports. Those reports outline important scientific results and highlights, a follow-up of quantitative indicators set by the programme proposals, and information on the budgetary. The reports of each field are used by president to inform the Senate about the progress of activities and any necessary readjustment.

HGF is responsible for planning and operating scientific large-scale facilities and infrastructures, like supercomputers, research ships and aircraft. The running and user management of infrastructures are independent for each centre. The life span and costs of the facilities is decided considering their long-term planning in the PoF proposals in a separate category from R&D activities. For their long-term reliability, HGF requires that all the infrastructures have at least €500 million running costs per year and at least 50% external users.

B4 Evaluation practices

The Executive Committee organises PoF evaluation in a process: composing the Senate Commission, collecting the Senate's recommendations on individual programmes then reaching final PoF decisions across the six fields. Numerous national and international research organisations are consulted for choosing reviewers based on candidates list proposed by the centres. A reminding note should be attached with the candidate list if any potential evaluator is inter-linked with the centre. The final evaluators have to keep in mind strategic and scientific focuses of HGF. The three days on-site PoF review includes presentations by scientific directors with strategic framework, by coordinators of the research field, by programme speakers about individual topics. Evaluators join the parallel sessions, discussing with SAB members and heads of the scientific councils participated. Schedules also include special sections about knowledge transfer, trainings involving early career researchers, and public engagement.

Centred on scientific quality and strategic relevance, reviewers give recommendations on programme and individual topics concerning the criteria as laid out in Table B4. The scientific excellence part (standing, originality, competence) has been mostly done by midterm reviews of the latest period. Large-scale restructures are evaluated every five years in parallel with the programme reviews and not uncompetitive with each other or R&D activities. Reviewers consider their access authorities, support for users, appropriateness of resources used and to be used. The resulting report is submitted to the Senate Commission and to president to make decisions. If the fixed financing proposal in the next five years does not exceed the expectation, some funding could be given back to the R&D part. There might be additional reviewers for large infrastructures, but most of the reviewers are the same.

At the end of each evaluation day, there are internal meetings with the Director. The closed session would end with some informal recommendations in an oral presentation of preliminary result. The formal report of one programme includes strategic significance and quality of programme topics addressing those criteria and recommendations. The overall seven-score rating ranges through *Outstanding–Excellent–Very good–Good–Average–Not sufficient*. The report is first submitted to the Senate Commission and the president. Then the negotiation between them and the centre goes on with further clarification, agreement and modification.

The Balancing Role of Evaluation Mechanisms

Evaluation criteria	For entire programmes	Standing	The international competitiveness of the programme and scientific competence of the contributing groups
		Composition	The selection and decision-making of suitable research themes, strategic partners and cooperation structures
		Potential	Talent management including diversity, recruitment processes, and access to suitable infrastructures, and overarching cross-programme activities and initiatives with other groups, universities and industry
		Added value	Transfer of knowledge and technology to stakeholders and ultimately into Society and providing access of external users to data and infrastructures
	For each individual topic within programmes	Originality	The originality and innovative potential of the future research plans, including the quality of implementation steps and milestones
		Competence	The overall scientific performance within given topics
		Alignment	Strategic relevance of Helmholtz mission, alignment with the Strategic Guidelines and Position Papers
		Coherence	Consistency of the programme and its topics completeness and integrity of relevant research themes, strategic partners and infrastructures
		Cooperation & management	The management of the topic and internal/external and national/international cooperation

Table B4 Evaluation criteria for the entire programme and each topic in programmes

Source: HGF headquarters interview, 2013

C CAS

The supplementary materials of the Chinese Academy of Sciences (CAS) were collected and interpreted from its official website (<http://english.cas.cn/>) and the published policy documents: 2013 annual report (CAS, 2013), Knowledge Innovation Programme (KIP) evaluation report (CAS, 2012), One–Three–Five evaluation report (Evaluation Centre report, 2014). This section also includes primary data (interview findings) and secondary literature sources.

C1 Historical dynamics

CAS has winding and complicated dynamics. It was established on November 1, 1949, just one month after the founding of the People’s Republic of China. Upon the establishment, magnificent domestic resources were invested in CAS to support its development. Since then, CAS has engaged in the national research agendas and has represented the national top research and development (R&D) institution and power house. Since the Open-Up Reform of China in 1978, CAS has expanded significantly in terms of scale, research quantity and quality. Over these years, CAS has expanded significantly and become the largest and most important R&D institution of China. The historical dynamics of CAS can be reflected by its evolving missions, strategies, and reforms in the past two decades, which are presented in the following sections.

C2 Organisational mission and structure

The missions of CAS are politically defined as to become the national academic leader in natural sciences R&D, a major advisory body to government S&T policies, and a national basement for incubating high technology in China's market. In July 2013, China's president Xi confirmed the triple roles of CAS—as research institutes, national academy and educational institution, to become the country's main source of innovative ideas, talent, and achievements. The *Four First* tasks were proposed then—first in achieving leap-forward development of S&T, first in building high grounds for cultivating innovative talent, first in establishing high-quality S&T national think tank, and first in building internationally leading research institutes. The title of individual membership of CAS, *Academicians*, represents a highly prestigious position for CAS scientists domestically.

In 2012, CAS took 58 projects of '973 Programme' (32% of the total), and 113 projects in '863 Programme' (10% of the total), both types funded by the Ministry of Science and Technology (MOST). Moreover, CAS conducted around 3,600 'General Project' (12% of the total), and 125 'Key Project' (21% of the total), both types funded by the National Natural Science Foundation of China (NSFC). There are other tens of thousands of projects funded by regional governments and industries.

As shown by Figure C2-1, in China's R&D landscape (S&T system usually called in China), CAS is positioned at the same political level as national ministries. There is strict top-down governance from the State Council to the CAS headquarters. All funding and reform plans of CAS have to be approved by the State Council. As discussed in section 4.1, MOST is an important partner of CAS to jointly accomplish national research programmes that are the major resources of competitive projects CAS institutes. The Ministry of Finance is in charge of public block funding of CAS. The Ministry of Education guides talent training in CAS universities.

Figure C2-1 China's R&D landscape

Source: Li et al., 2008

CAS has a powerful headquarters, as shown by Executive Committee, Bureaus and Academic Divisions combined in Figure C2-2. This figure represented the CAS organisational structure in 2013, but a new structure after the reform of Pioneer Action Plan (PAP) has not been published. As presented in section 4.2, CAS has 13 regional branches, 104 research institutes, 3 universities with 1 graduate school, 4 science libraries, 3 technology support centres, 2 news and publishing units, and over 430 S&T-based enterprises in 11 industries, including 8 companies listed on stock exchanges. All of these units are located throughout China. Nine provincial branches as regional administration offices participate in some important issues of CAS institutes, such as human resource management, appointment of new directors, and collaborations with local governmental research organisations and industries as a communication hub between L1 and L2. In addition, as an important educational base for high-level talent of China, the three CAS universities, namely, University of Chinese Academy of Sciences (UCAS), University of Science and

Technology of China (USTC), and the newly developing Shanghai University of Science and Technology, mainly foster graduate students. In 2012, the three universities awarded 6,400 master and 5,600 PhD degree graduates.

Figure C2-2 Organisational structure of CAS

Source: CAS annual report, 2013

Attracting talent from abroad has been promoted by a series of CAS policies. As a significant start, the CAS '100 Talent Plan' started in 1994 with high-level targets, recruiting 100 outstanding scholars from both home and abroad by the end of the 20th century, and with intensive support from public funding. Over the past two decades, CAS has recruited more than 1000 distinguished scientists from abroad. In 2015, this programme was redefined aiming at three specific categories of talent: academic director, engineering director, and young talent. Some followed-by policies have been issued, such as the '1,000 Talent Plan' since 2010 and '10,000 Talent Plan' since 2012, international fellowships, and public innovation platforms with international standard. The number of foreign researchers in CAS is expected to rise from 1% in 2013 up to 3% by 2020.

C3 Reforms

The following descriptions of the five reforms supplement CAS's key strategy of *Major Outcome Orientation* proposed in section 4.3.1. The Major Outcome Orientation was proposed in the 2011 National S&T Innovation Conference to discuss the objectives, priorities, and tasks of the national reform for *Innovation 2020*. Major Outcome from CAS institutes was defined in accordance with six Academic Divisions. The definition and the relevant standards and requirements in each Division were specified by examples, which can be summarised as: overcoming a big problem existing in one field for a long time; creating a new research direction, new discovery, theory, or interpretative framework; breakthrough in key technologies; solutions for systemic problems; technology transfer with major societal or economic effects; and significant consultancies.

First, KIP from 1998 to 2010 was the forerunner to categorise and restructure the large number of CAS institutes. There are a series of consequences: 10 prioritised funding fields (e.g. health) were created; the number of legally independent institutes was reduced from 123 to 99; 46 institutes were reconstructed to 16; 12 institutes were newly established; 7 institutes' research direction totally changed; and 6 institutes turned into research spin-offs (KIP evaluation report, 2012). About half of CAS researchers lost their permanent positions to cut the costs of maintaining the massive research community of CAS (Liu et al., 2011). These changes resulted in a preliminary formation of CAS research structure which aimed at aligning with international standards of research management such as resource allocation. Collaboration across disciplines and institutes was strongly encouraged. KIP was proved by many scholars to lead to CAS's increased research quantity and quality and to narrow the R&D gap between China and developed countries. The improvement of scientific structure and management after KIP led to the increase of productivity of CAS institutes by 12.5% among which 8.8% was attributed to technological progress and 3.3% to efficiency improvement (Zhang et al., 2011). KIP also promoted recruitment of leading talent from abroad and domestic universities (Liu et al., 2011). High salaries, generous financial support and prestige position are considered as key factors to reduce the brain drain of scientists for China (Liu et al., 2011; Suttmeier et al., 2006).

Second, *Innovation 2020* emphasised indigenous innovation of institutes by another portfolio of stimulating policies based on KIP. Its 16 Major Programmes cultivated over 2000 outstanding S&T talent, invested costly but high-quality

research facilities for international competitiveness and reputation, and transferred over 120,000 leading young talent from universities to industries. The *Nice Shifts in Innovation 2020* restructured the units guided by distinct categories of research orientations. 17 new research fields were proposed to map within the *Innovation 2050: China's S&T future*.

Third, a substantial reaction from headquarter was its restructuring internal administrative Bureaus in 2013. Previous Bureaus, originally as administration centres for individual institutes, had overlapping functions and lacked efficient coordination. New Bureaus of *Fundamental research*, *Major R&D Programmes*, *S&T Development* connecting with ministries, *Frontier Science and Education* linking to universities, *Strategic Planning* and *Finance and Support* dealing with internal affairs, were better classified according to research orientations and guided relevant administrations more efficient.

Fourth, CAS's mission has adopted a *Three-in-One* principle to systemise its three distinct roles: policy support as a national academy, science exploration as research institutes, and education as universities. With the help of newly categorised administrative Bureaus in headquarters, the three roles of each institute are clearly positioned and collaboration among institutes concerning each role is strengthened. Therefore, activities of an institute in relation to a certain role are managed or coordinated by one particular Bureau, rather than the previous overlapping management of several Bureaus.

Fifth, PAP was proposed in the 2014 CAS Congress as the pioneer of structural reforms of CAS institutes. The newly established research units have been categorised into the four distinct research orientations according to their nature and goals of the current and future research activities (Poo & Wang, 2015). Each unit will be equipped with integrated resources and talent of the entire CAS.

The first category is *Centres of Excellence* focusing on basic research in the frontier sciences, and the areas where China has a chance to dominate. These centres put together high level expertise selected from various institutes to solve great challenges, develop new research areas, and achieve breakthroughs with major impact worldwide. For instance, one team selected 40 researchers from 11 different CAS institutes specialising in brain science for neurodegenerative-disease research. So far 4 centres have been established in the areas of quantum information, Tibet Plateau Geo-system, particle physics, and brain science. There is a target of 20 centres by 2020.

The second category *Innovation Academies* aims at larger-scale applied R&D outputs in technological and industrial areas with underdeveloped commercial potentials. Creating such a academy has to synergise systemically the strengths of technology transfer of the existing institutes working in one field rather than just split those applicable units from the institutes and then put them together in one academy. 5 academies have been designed at the initial phase, each working in a particular field, such as information technology and medicine R&D. There is a target of 20 innovation academies by 2020.

Big Science Facilities, the third group, aim to provide major R&D facilities and platforms that are open domestically and internationally. It promotes collaborators using the same facilities for both fundamental studies and industrious research at different regions. Generally, CAS has about 70% of all R&D facilities in China which brings about cooperation strength. The conceptual systemising those big-scale and expensive facilities, located at different regions, refers to unified utilisation, management and coordination rules considering geographic boundaries in order to improve the integral efficiency.

The fourth category *Specialty Institutes* are devoted to research in specialised areas relevant to regional development and people's life, such as preparation for natural disasters. The examples are Institute of Mountain Hazards and Environment in Chengdu, and Institute of Psychology. This group aims to assist local development and sustainability and is still in the design phase.

In total about 70 institutes can be included in these four categories. Each new unit in the four categories consists of researchers across the previous institutes and centres. The restructuring aims at integration of research goals and stimulating collaboration and teamwork across groups in each category. Some traditional institutes established on the basis of disciplines such as Institutes of Chemistry, Physics, etc. will become department partners of CAS universities to integrate research and education in particular disciplines, and also to keep those traditional disciplinary institutes independent. Other institutes and centres that do not fit into the above scheme will be merged.

C4 Evaluation practices

This section supplements the thesis section 6.3.1 about evaluation institutionalisation of CAS. The concept of S&T evaluation came into the spotlight of CAS in the 1990s when it was transforming out of the completely planned system. To create research evaluation for efficient resource allocation became a first step. Since then, CAS has experienced a series of institutional reforms which use evaluation as a key notion and an important tool of the reforms. After KIP, extra resources beyond fixed income started to be allocated to institutes which are proved to perform superiorly by evaluation. Evaluation Centre of CAS was established in 2000 as a claimed independent support centre of the headquarters. Since then systemic evaluation towards all the institutes started to be planned and organise the Evaluation Centre.

Horizontal programmes and projects follow particular assessment procedures required by funding bodies. Longitudinal projects usually comply with standardised procedures: *ex-ante* proposal approval, midterm assessment and final check, by MOST, or NSFC, or CAS itself. Those procedures required by multiple sources (see Figure C2-2) are coordinated by Department of Research Management of each institute. In the coordination processes, the potentiality of aggregating scattered projects is often ignored, which leads to short-term behaviours, such as repetitious and fragmented projects further exacerbated by valuing the success of previous grant applications (interview findings).

The attraction and retention of human resources involve talent evaluation in two ways: talent attraction programme with specific recruitment procedures (such as 1,000 Talent Plan and 10,000 Talent Plan) and annual evaluation of senior and junior researchers by the administration office of each institute. The evaluation of laboratories and their infrastructures takes another system, independent from institute evaluation, related to regular operation and is an input-output assessment of infrastructures for ranking of and awards among laboratories.

About the unit responsible for the evaluation mechanism of CAS, there was only one Evaluation Unit in the Planning and Development Bureau with less than ten persons in the 1990s. In 2000, the Evaluation Centre was established particularly for the KIP as an independent support centre with about thirty staff. It works in the Institute of Policy and Management (IPM) in Beijing, a unique institute in CAS specialising in social impact of S&T policy, management, and evaluation.

Before 2012, the three main types of evaluation in CAS were shown by Table C4. Because the One–Three–Five evaluation has not yet permeated into each institute, Table C4 still helps to understand the existing CAS evaluation mechanism.

Types	Subdivided	Evaluation grading
Binary self-evaluation: supporting strategic management	Situation analysis of CAS	
	Assessment of S&T layout	
	Plan and policy evaluation	
	Evaluation of S&T Innovation Base	
Regular institute evaluation: to promote competition and development	Overall quality evaluation	
	Innovation capacity Monitoring	
	Input–output analysis of institutes	
Specific (ad hoc) evaluation: demonstrating policy orientation	Project evaluation	
	Evaluation of human resource	
	Evaluation of Laboratory	
	Assessment of policy guidance	

Table C4 The three key evaluation types of CAS before 2012

Source: Li et al., 2008

The three-level indicator system applied in 2007 (section 6.4.1) included qualitative peer review at the top level and annual statistics as the bases (talent quantities and funding per capital and awards). The middle-level key indicators for individual institutes and their comparisons involved research projects and programmes, invitation for international conferences and workshops, construction and operation of research facilities, platform of public resources, intellectual property, high-quality papers, engineering application and demonstration, and advisory reports. However, limited numbers of indicators were selected: only six key indicators of each institute, three common ones and eight specific ones for institutes' comparison. Afterwards, dimensions for innovative capacity of institutes were increased to 5 aspects and 24 indicators, as shown in Figure C4. Meanwhile, simple benchmarks began to reflect the shared characteristics between institutes. This indicator-dependent model was operationalised with great efforts. Based on that, experts' reviews would check qualitative research and management goals by assessing individual institutes and their synergies. In practice, only limited numbers of international peer-review were undertaken in CAS institutes.

Practice 2: Quantitative Monitoring on Innovation Capacity

Based on 24 fundamental indicators

Figure C4 The three-level evaluation indicator system of CAS

Source: Li et al., 2008

The *One–Three–Five evaluation* (*one position, three breakthroughs and five key fostering directions*) towards the target of Major Outcome of CAS institutes has been applied since 2012. The definition of Major Outcome in each Academic Division and research field experienced diverse discussions among CAS heads, scientists, and international experts. The specific data and rules about *One–Three–Five evaluation* are confidential, such as the mission statement signed by each legally independent institute. Only institute directors have access to all the relevant documents for decision-making and necessary adjustment. During each round of *One–Three–Five evaluation*, the specific *One–Three–Five aspects* are initiated by institutes themselves following the requirements from headquarters. The changes of research direction and financial focuses should be described in details in the annual report for the final check in five-year period. The relevant adjustment to those changes is also prepared by institutes and reported to headquarters. The feedback mechanism about how to communicate the final detailed evaluation results and the related adjustment decisions was uncertain when this study took interviews in CAS. Should the evaluation results and uses be negotiated only between the president and institute directors, or also involving CAS executive committee, relevant Bureaus, and the Evaluation Centre? These can be explored in future research. The following boxes show examples of responses from several CAS institutes towards the *One–Three–Five* evaluation in 2013.

Box C4–1: Responses from CAS institute a

Our institute was chosen as one of the first pilot institutes to implement the One–Three–Five evaluation in 2014. Our institute ranks in the top three of CAS institutes regarding domestic and international collaboration. These experiences become the rich resources for constructing international review panels in the One–Three–Five evaluation. The potential candidates come from the institute itself, which knows best who are the real experts in its domain, both nationally and internationally. The institute’s proposal that provided for the One–Three–Five aspects focuses on the clear position of the institute and resources distributed for the 12 innovation teams in our institute. The three breakthroughs and five directions represent the future orientation. There is still some other research through external projects beyond the One–Three–Five aspects. We can see at this moment, the One–Three–Five evaluation aims to promote the Major Outcome that is required by the State Council. After that, there will be more annual stable funding, which used to be competitive grants, from the headquarters to our institute. The increased funding is expected to be used for the three breakthroughs and five directions, and also for creating excellent research projects to attract talent.

Box C4–2: Responses from CAS institute b

Our institute is large-scale and complex, the previous evaluations tried to compare and even rank research outcomes of various PI-led research groups, and of different research programmes, which is so difficult and people hate that. Neither the high impact factors of publications in journal papers nor the number of patent should become the only assessment criteria. Some R&D in our institute focuses on engineering and construction in poor and remote suburbs in China, which is also breakthrough and cannot be compared with publications. The One–Three–Five evaluation facilitates us to exchange feedback and learn experiences among heterogeneous groups or programmes. Now we choose the way of public presentations at the annual assembly to discuss research quality of different groups or programmes, outcomes, potentials, and impacts to all the audiences. There will be no rating, ranking or comparison. The increasing communication inside the institute aims to aggregate the breakthroughs and key fostering directions of many and heterogeneous groups. However, it is at the early stage. There are also some debates between groups on selecting the most representative breakthroughs and future developments of the whole institute.

Box C4–3: Responses from CAS institute d

There are different definitions and standards for the three breakthroughs and five key directions in our institute; see examples, the most important national or even international S&T awards; publications in famous journal, such as Nature, Science, and Cell; impressive achievement in particular practical R&D fields. Before signing the mission statement, we prepared specific research plans and also indicators for assessment in relation to the One–Three–Five aspects to be achieved in the next five years. Only if that plan was accepted by the Academic Division of headquarters, our institute would sign the mission statement and meanwhile receive part of the funding to conduct the One–Three–Five–aspects–related R&D activities. During the next five-year period, we should prepare annual reports highlighting the achievements and progress of the One–Three–Five aspects we proposed. This way is more simple and efficient compared to the previous administrative annual reports. In five years, the final check will be made even we have a new director then. This One–Three–Five evaluation is still at a very early stage. The pilot projects select large-scale institutes and aim to take advantage of their experiences of international evaluation. Our institute prepared for the One–Three–Five aspects in 2010 and signed the mission statement in 2011. The first check of our achievement of the One–Three–Five aspects in the mission statement will take place in 2015.

References

- AeA. (2007). China's 15 year science and technology plan. *The AeA Competitiveness Series*, 14, 4.
- Albaek, E. (1995). Between knowledge and power—utilization of social science in public policy making. *Policy Sciences*, 28(1), 79–100.
- Amanatidou, E., Cunningham, P., Gök, A., & Garefi, I. (2014). Using evaluation research as a means for policy analysis in a 'new' mission-oriented policy context. *Minerva*, 52(4), 419–438.
- Argyris, C. (1977). Double loop learning in organisations. *Harvard Business Review*, 55(5), 115–125.
- Arnold, E (2003). Evaluating research and innovation policy: a systems world needs systems evaluations, *Research Evaluation*, 12 (3), 3–17.
- Arundel, A., Lorenz, E., Lundvall, B. Å., & Valeyre, A. (2007). How Europes economies learn: a comparison of work organisation and innovation mode for the EU–15. *Industrial and Corporate Change*, 16(6), 1175–1210.
- Asheim, B. T., Boschma, R., & Cooke, P. (2011). Constructing regional advantage: platform policies based on related variety and differentiated knowledge bases. *Regional Studies*, 45(7), 893–904.
- Asheim, B. T., & Coenen, L. (2005). Contextualising regional innovation systems in a globalising learning economy: on knowledge bases and institutional frameworks. *Circle Lund Univesity*, 163–173.
- Bai, C. (2007). Some considerations on science and technology evaluation of CAS. *Internal Journal of CAS*, 22 (2), special report from CAS S&T evaluation research group, 22 (2), 104–114.
- Bai, C. (2012). Reform of CAS S&T evaluation: towards a major R&D outcome-orientated system. *Internal Journal of CAS*, 27 (4), special issue *Innovation 2020*, 407–410.
- Bamberger, M. (1991). The politics of evaluation in developing—countries. *Evaluation and Programme Planning*, 14(4), 325–339.
- Baumann, O. (2015). Models of complex adaptive systems in strategy and organisation research. *Mind Soc*, 14, 169–183.
- Beck, U., Bonss, W., & Lau, C. (2003). The theory of reflexive modernization—problematic, propositions and research programme. *Theory Culture & Society*, 20(2), 1–33.
- Benz, A. (1997). Organisation and network—institutional control in economics and politics. *Politische Vierteljahresschrift*, 38(4), 877–878.

- Biber–Klemm, S. (2011). *Responsibilities of stakeholders in academic research*. International workshop 'Access to genetic resources and the sharing of benefits arising from their utilization', Swiss Academy of Sciences, Bern.
- Block, H. J., & Krull, W. (1990). What are the consequences—reflections on the impact of evaluations conducted by a science policy advisory body. *Scientometrics*, 19(5–6), 427–437. doi: 10.1007/bf02020705
- Bohren, O., Josefsen, M. G., & Steen, P. E. (2012). Stakeholder conflicts and dividend policy. *Journal of Banking & Finance*, 36(10), 2852–2864.
- Borrás, S. (2012). *Three tensions in the governance of science and technology*. The Oxford Handbook of Governance.
- Borrás, S., & Edler, J. (2014). Introduction: on governance, systems and change, handbook of the governance of socio–technical systems, *Explaining Change*. Edward Elgar: EU–SPRI, 1–22.
- Borrás, S., & Højlund, S. (2014). Evaluation and policy learning: the learners' perspective. *European Journal of Political Research*, 54(1), 99–120.
- Boswell, C. (2008). The political functions of expert knowledge: knowledge and legitimation in European Union immigration policy. *Journal of European Public Policy*, 15(4), 471–488.
- Botcheva, L., et al. (2002). Learning culture and outcomes measurement practices in community agencies. *American Journal of Evaluation*, 23(4), 421–434.
- Braun, D. (2003). Lasting tensions in research policy-making—a delegation problem. *Science and Public Policy*, 30(5), 309–321.
- Braun, D., & Guston, D. H. (2003). Principal–agent theory and research policy: an introduction. *Science and Public Policy*, 30(5), 302–308.
- Bryson, J. M., Patton, M. Q., & Bowman, R. A. (2011). Working with evaluation stakeholders: a rationale, step-wise approach and toolkit. *Evaluation and Programme Planning*, 34(1), 1–12.
- Byung, H., Byeon, S., Byung, Y. (2010). The Influence of the government–funded research institutes (GRIs) evaluation system on the acceptability of evaluation result and performance of the GRIs. *Asian Research Policy*, 1(2010), 120–136.

- Bussmann, W. (1996). Democracy and evaluation's contribution to negotiation, empowerment and information, some findings from Swiss democratic experience. *Evaluation* 2(3): 307–319.
- Campbell, D. (2003). *The evaluation of university research in the United Kingdom and the Netherlands, Germany and Austria*, in Shapira P. and Kuhlmann S. (eds.): Learning from science and technology policy evaluation. experiences from the United States and Europe, Cheltenham: Edward Elgar, 98–131.
- Cao, J. (2012). Presentation about science technology and innovation in China. Amsterdam, Netherlands.
- CAS. (2012). *KIP evaluation report: experiences of China's national innovation system*. Beijing, China: Science Press.
- CAS. (2014). Annual report. Retrieved 2014, from http://english.cas.cn/about_us/reports/
- CAS Evaluation Centre. (2013). *An introduction to expert diagnosis assessments for CAS institutes*. Unpublished report, CAS, Beijing, China.
- Ceci, S. J., Williams, W. M., & Mueller–Johnson, K. (2006). Is tenure justified? An experimental study of faculty beliefs about tenure, promotion, and academic freedom. *Behavioral and Brain Sciences*, 29(6), 569–594.
- Cicchetti, D. V. (1991). The reliability of peer-review for manuscript and grant submissions—a cross-disciplinary investigation. *Behavioral and Brain Sciences*, 14(1), 119–134.
- Coglianese, C. (2012, August). Measuring regulatory performance: evaluating the impact of regulation and regulatory policy, OECD, *Expert Paper*, pp. 1–59.
- Cozzens, S. (1990). *Options for the future of research evaluation*, in Cozzens S., Healey P., Rip A., and Ziman J. (eds.): The Research System in Transition, Dordrecht: Kluwer, 281–294.
- Cozzens, S., (2003). *Frameworks for evaluating S&T policy in the United States*, In: Shapira, P. and Kuhlmann, S. (eds.): Learning from Science and Technology Policy Evaluation, Elgar, 54–64.
- Cozzens, S. (2007). *Death by peer review*. In R. Whitley & J. Gläser. (eds.): The Changing Governance of the Sciences, Springer, 225–242.
- Crow, M., & Bozeman, B. (1998). *Limited by design*: Colombia University Press.
- Cruz–Castro, L., & Sanz–Menendez, L. (2007). *Research evaluation in transition: individual versus organisational assessment in Spain*. In R. Whitley & J. Gläser (eds.): The Changing Governance of the Sciences, Springer, 205–223.

- Cuba, E., & Lincoln, Y. (1989). *Fourthe generation evaluation*. USA: Sage Publications.
- Cyranoski, D. (2014). Chinese science gets mass transformation. *Nature*, 513 (7519), 468–469.
- Daft, R. *Organisational theory and design* (2010). 9th Edition. Mason, OH: Thomson.
- Dahler–Larsen, P. (2011). *Evaluation Society*. Stanford University Press.
- Dai, T., Xiao, X., & Li, X. (2014, October). *The Context-based selection of evaluation models: the case of Chinese Academy of Sciences (CAS)*. Paper presented at the the 11th European Evaluation Society (EES) Biennial Conference, Dublin, Ireland.
- Dimaggio, P. J., & Powell, W. W. (1983). The iron cage revisited—institutional isomorphism and collective rationality in organisational fields. *American Sociological Review*, 48(2), 147–160.
- DiMaggio, P. J. and W. W. Powell (1991). *The new institutionalism in organisational analysis*. The University of Chicago Press.
- Dorbeck–Jung, B., & Shelley–Egan, C. (2013). Meta–regulation and nanotechnologies: the challenge of responsabilisation within the European Commission's code of conduct for responsible nanosciences and nanotechnologies Research. *Nanoethics*, 7(1), 55–68.
- Edelenbos, J., & Van Buuren, A. (2005). The learning evaluations: a theoretical and empirical exploration. *Evaluation Review*, 29(6), 591–612.
- Eden, C. & Ackermann, F. (1998) *Makina strategy: the journey of strateaic management*. London: Sage Publications.
- Edler, J., Berger, M., Dinges, M., & Goek, A. (2012). The practice of evaluation in innovation policy in Europe. *Research Evaluation*, 21(3), 167–182.
- Edquist, C. (1997). *Systems of Innovation, Technologies, Institutions, and Organisations*. London and Washington Press.
- Edquist, C., & Hommen, L. (1999). Systems of innovation: theory and policy for the demand side. *Technology in Society*, 21(1), 63–79.
- EFI (Expertenkommission Forschung und Innovation) report. (2012). *Research innovation and technological performance in Germany*. Berlin, Germany.
- EFI report (2015). *Research, innovation and technological performance in Germany*. Berlin, Germany: EFI.
- Fang, X (2012). *KIP evaluation report of 1998–2010, experiences of China's national innovation system*, CAS, Science Publisher, Beijing, China.

- Fang, Y., & Tian, D. (2012). *Report of China's science and technology evaluation system*. Beijing, China: MOST.
- Feller, I. (2007). Mapping the frontiers of evaluation of public-sector R&D programs. *Science and Public Policy*, 34(10), 681–690.
- Fetterman, D. (1996). *Empowerment evaluation: an introduction to theory and practice*. In D. Fetterman, S. Kaftarian & A. Wandersman (eds.): Empowerment Evaluation Knowledge Tools for Self-Assessment & Accountability. USA: Sage, 1–18.
- Freeman, C. (1987). *Technology and economic performance: lessons from Japan*. Pinter, London.
- Freeman, C. (1995). The national system of innovation in historical-perspective. *Cambridge Journal of Economics*, 19(1), 5–24.
- Freeman, C. (2002). Continental, national and sub-national innovation systems—Complementarity and economic growth. *Research Policy*, 31(2), 191–211.
- Freeman, R. E. (1984). *Strategic management: a stakeholder approach*. Boston, MA: Pinnar
- Fisher, E. & Schuurbiers, D. (2013). *Socio-technical integration research: collaborative inquiry at the midstream of research and development*. In: Doorn, N. et al (eds.): Early engagement and new technologies: Opening up the laboratory. series philosophy of engineering and technology, 16, 97–110.
- Fuchs, L. S., Fuchs, D., Hamlett, C. L., Walz, L., & Germann, G. (1993). Formative evaluation of academic progress—how much growth can we expect. *School Psychology Review*, 22(1), 27–48.
- Fuchs, S. and J. H. Turner. (1986). What makes a science “mature”? patterns of organisational control in scientific production, *Sociological Theory*, 4, 143–150.
- Fukuyama, F. (1995). Social capital and the global economy. *Foreign Affairs*, 74(5), 89–103.
- Gazlig, T (2009). *The strategy of the Helmholtz Association: top-level research for society, science and the economy*. HGF, Berlin, Germany.
- Gazlig, T (2012). *Helmholtz-roadmap for research infrastructures*. HGF, Berlin, Germany.
- George, A., & Bennett, A. (2005). *Case studies and theory development in the social sciences*. Harvard University.

- Gläser, J, and Laudel, G (2007). *Evaluation without evaluators: the impact of funding formula on Australian university research*, in Whitley and Gläser (eds.): *The changing governance of the sciences: the advent of research evaluation systems*, 127–149.
- Golden-Biddle, K., & Rao, H. (1997). Breaches in the boardroom: organisational identity and conflicts of commitment in a nonprofit organisation. *Organisation Science*, 8(6), 593–611.
- Good, B., Vermeulen, N., Tiefenthaler, B., & Arnold, E. (2015). Counting quality? The Czech performance-based research funding system. *Research Evaluation*, 24(2), 91–105.
- Georghiou, L. (1995). Research evaluation in European national science and technology systems, *Research Evaluation*, 5(1), 3–10.
- Georghiou, L., (2003). *Evaluation of research and innovation policy in Europe – new policies, new frameworks?* In: Shapira, P. and Kuhlmann, S (eds.): *Learning from Science and Technology Policy Evaluation*, Elgar, 65–80.
- Georghiou, L. and Laredo, P. (2006), *Evaluation of publicly funded research. recent trends and perspectives*. OECD [DSTI/STP (2006)7], Paris.
- Georghiou, L., & Kuhlmann, S. (2002). *Future Policy Instruments: Evaluation of the Socio–Economic Effects in the European Research Area*. In: Polt. W. / Rojo, J. / Zinöcker, K. / Fahrenkrog, G. / Tübke, A. (eds.): *RTD Evaluation Tool Box.—Assessing the Socio–Economic Impact of RTD–Policies Brussels/Luxembourg* (EU, Joint Research Centre, Institute for Prospective Technological Studies, CSC–EEC–EAEC) (IPTs Technical Report Series, EUR 20382 EN), 203–210.
- Geuna, A., & Martin, B. (2003). University Research Evaluation and Funding: An international comparison. *Minerva*, 41, 277–304.
- Grayson, M. (2015). China’s research under the magnifying glass. *Nature Outlook*, 520, 1–28.
- Guerrieri, P, & A. Tylecote (1997). *Interindustry differences in technical change and national patterns of technological accumulation*, in Edquist, C (ed.), *Systems of Innovation: Technologies, Institutions and Organisations*. London and Washington Pinter, 107–125.
- Hage, J., Jordan, G., & Mote, J. (2007). A theory-based innovation systems framework for evaluating diverse portfolios of research, part two: macro

- indicators and policy interventions. *Science and Public Policy*, 34(10), 731–741.
- Hansen, F. (2013). Systemic evaluation governance: new logics in the development of organisational fields. *Scandinavian Journal of Public Administration*, 16, 47–64.
- Hansson, F. (2006). Organisational use of evaluations: governance and control in research evaluation. *Evaluation*, 12(2), 159–178.
- Hardwig, J. (1991). The Role of Trust in Knowledge. *Journal of Philosophy*, 88(12), 693–708.
- Hart, P. & Kleiboer, M. (1995). Policy controversies in the negotiatory state, Knowledge and Policy: *The International Journal of Knowledge Transfer and Utilization*, 8(4), 5–25.
- Hartmann, I., & Neidhardt, F. (1990). Peer-review at The Deutsche-Forschungsgemeinschaft. *Scientometrics*, 19(5–6), 419–425.
- Heinze, T., & Kuhlmann, S. (2008). Across institutional boundaries? research collaboration in German public sector nanoscience. *Research Policy*, 37(5), 888–899.
- Henry, G. T., & Mark, M. M. (2003). Beyond use: understanding evaluation's influence on attitudes and actions. *American Journal of Evaluation*, 24(3), 293–314.
- Hermanson, D. R., & Rittenberg L. E. (2003). *Internal audit and organisational governance*, in Bailey, A.D., Audrey A., and Sridhar R., (eds.): Research opportunities in internal auditing. Altamonte Springs, FL: The Institute of Internal Auditors Research Foundation, 25–71.
- HGF. (2014). *Annual Report 2013: Helmholtz-Research for Change*. Retrieved 2014, from http://www.helmholtz.de/no_cache/en/about_us/press_and_media/media_centre/publications/pr/annual-report-2013-helmholtz-research-for-change-1/
- Hodgson, G. M. (1993). Theories of economic evolution—a preliminary taxonomy. *Manchester School of Economic and Social Studies*, 61(2), 125–143.
- Hofstede, G.H. (1991), *Cultures and Organisations: Software of The Mind*. London; New York, McGraw–Hill Book Company.
- Højlund, S. (2014). Evaluation use in the organisational context—changing focus to improve theory. *Evaluation*, 20(1), 26–43.

- Hood, C. (1995). The "New Public Management" in the 1980s: variations on a theme. *Accounting, organisations and society*, 20(2), 93–109.
- Hoppe, R. (2011). Institutional constraints and practical problems in deliberative and participatory policy making. *Policy & politics*, 39(2), 163–186.
- Hoppe, R. (2009). *Governance of Problems. Puzzling, Powering, and Participation*. Policy Press, Bristol.
- HRK. (2011). *Higher Education Finance*. Retrieved October, 2015, from <http://www.hrk.de/activities/higher-education-finance/>
- Hutchison, J. R. (1995). A multimethod analysis of knowledge use in social policy: research use in decision affecting the welfare of children. *Science Communication*, pp. 90-106.
- Irvine, J., & Martin, B., (1984). *Foresight in Science: Picking the Winners*, Frances Printer: London.
- Jacob, S., Speer, S., & Furubo, J. E. (2015). The institutionalisation of evaluation matters: updating the international atlas of evaluation 10 years later. *Evaluation*, 21(1), 6–31.
- Jansen, D. (2007). *New Forms of Governance in Research Organisations. Disciplinary Approaches, Interfaces and Integration*. Dordrecht: Springer.
- Jasanoff, S. (1989). Norms for evaluation regulatory science. *Risk Analysis*, 9(3), 271–273.
- Jasanoff, S. (1995). Procedural choices in regulatory Science. *Technology in Society*, 17(3), 279–293.
- Jasanoff, S (2007). *Designs on nature: science and democracy in Europe and the United States*. Princeton University Press.
- Jepperson, R., & Meyer, J. W. (2011). Multiple levels of analysis and the limitations of methodological individualisms. *Sociological Theory*, 29(1), 54–73.
- Johnson, B. (2011). From user–producer relations to the learning economy. *Science and Public Policy*, 38(9), 703–711.
- Johnson, B (1992). *Organisational learning*, in Lundvall (ed.), *National Systems of Innovation, towards a theory of innovation and interactive learning*. London and New York Printer, 23–42.
- Johnson, B (1997). *Institutions and organisations in systems of innovation*, in Edquist, C (ed.), *Systems of Innovation Technologies, Institutions and Organisations*, London: Pinter Publishers/Cassell Academic, 41–60.

- Krull, W. (1995). The Max Planck experience of evaluation. *Scientometrics*, 34(3), 441–450.
- Kuhlmann, S., & Edler, J. (1998). *Clearing House*. Paper presented at the Helsinki Institute of Science and Technology Studies Helsinki, Finland.
- Kuhlmann, S. (1999). Moderation of policy-making? Science and technology policy evaluation beyond impact measurement —the case of Germany. *Evaluation*, 4(2), 130–148.
- Kuhlmann, S. (2003). Evaluation of research and innovation policies: a discussion of trends with examples from Germany, *Int. J. Technology Management*, 26, Nos. 2/3/4, 131–149.
- Kuhlmann, S. & Heinze, T. (2004): Evaluation von Forschungsleistungen in Deutschland: Erzeuger und Bedarf; Teil I: Konzeptionelle Grundlagen. *Wissenschaftsrecht*, 37 (2004), 53–69.
- Kuhlmann, S. & Heinze, T. (2004): Evaluation von Forschungsleistungen in Deutschland: Erzeuger und Bedarf; Teil II: Produktion und Verwendung evaluativer Information sowie Möglichkeiten ihrer künftigen Organisation. *Wissenschaftsrecht*, 37 (2004), 125–149.
- Kuhlmann, S. (2015). *Beyond crude measurement of evidence: evaluation as medium for learning*. Presentation at Jahrestagung der DeGEval—Gesellschaft für Evaluation, Speyer, Germany.
- Kuhlmann, S., et. al (2015): *Responsibility navigator*, Karlsruhe/Germany (Fraunhofer ISI), 132-155. Retrieved December 2015, from www.responsibility-navigator.eu.
- Kupferschmidt, K., & Vogel, G. (2014). Doing the math in Berlin. *Science*, 344(6186), 791–792.
- Lam, T. C. M. (1995). Individualization of objectives: a model of educational programme evaluation. *Evaluation Practice*, 16(1), 13–27.
- Laudel, G., & Gläser, J. (2014). Beyond breakthrough research: epistemic properties of research and their consequences for research funding. *Research Policy*, 43(7), 1204–1216.
- Laudel, G., & Gläser, J. (2006). Tensions between evaluations and communication practices. *Journal of Higher Education Policy and Management*, 28(3), 289–295.
- Leisyte, L., Enders, J., & de Boer, H. (2009). The balance between teaching and research in Dutch and English universities in the context of university governance reforms. *Higher Education*, 58(5), 619–635.

- Leviton, L. C., & Hughes, E. F. X. (1981). Research on the Utilization of Evaluation—A Review and Synthesis. *Evaluation Review*, 5(4), 525–548.
- Li, X., Shi, B., & Zhou, J. (2009). *A study of the model and methodology for institute evaluation in CAS*. Paper presented at the Evaluation Conference Orlando, Florida, USA.
- Liu, F., Simon, D. F., Sun, Y., & Cao, C. (2011). China's innovation policies: evolution, institutional structure, and trajectory. *Research Policy*, 40(7), 917–931.
- Lundvall, B. Å. (1992). *National systems of innovation: towards a theory of innovation and interactive learning*. London: Printer.
- Lundvall, B. Å. (2007). National innovation systems—analytical concept and development tool. *Industry and Innovation*, 14(1), 95–119.
- Lundvall, B. Å., Johnson, B., Andersen, E. S., & Dalum, B. (2002). National systems of production, innovation and competence building. *Research Policy*, 31(2), 213–231.
- Luo, J., Ordóñez–Matamoros, G., & Kuhlmann, S. (2015). Aggregated governance by R&D evaluation mechanism—case study of Chinese Academy of Sciences. *Asian Research Policy*, 6(2015), 56–72.
- March, J. G. (1991). Exploration and exploitation in organisational learning. *Organisation Science*, 2(1), 71–87.
- Mark, M. M., Henry, G. T., & Julnes, G. (1999). Towards an integrative framework for evaluation practice. *American Journal of Evaluation*, 20(2), 177–198.
- McCarthy, S. (2011). *How to write a competitive proposal for Horizon 2020*. McCarthy Hyperion Ltd.
- Meyer, J. W., & Rowan, B. (1977). Institutionalised organisations – formal–structure as myth and ceremony. *American Journal of Sociology*, 83(2), 340–363.
- Michelle, G. (2015). Raising standards. *Nature*, 520, 13.
- Mintzberg H., (1994). *Rise and Fall of Strategic Planning*, New York: The Free Press.
- MPG (2009). *Rules for Scientific Advisory Board*. Retrieved 2013, from <https://www.mpg.de/197429/rulesScientificAdvisoryBoards.pdf>
- MPG (2013). *Annual reports of the Max Planck Society*, Retrieved 2013, from <https://www.mpg.de/annual-report>
- MPG (2015). *Rules for Scientific Advisory Boards*, Retrieved 2015, from <https://www.mpg.de/197429/rulesScientificAdvisoryBoards.pdf>
- Nature Publishing Index (2014). Top 200 institutions. *Nature* 515, S98–S108.
- Ni, X. (2015). China's Research & Development Spend. *Nature* 520, S8–S9.

- North, D. C. (1991). Institutions. *Journal of Economic Perspectives*, 5(1), 97–112.
- OMB (2005). *R&D Investment Criteria*. The US Office of Management and Budget. Retrieved 2015, from <http://energy.gov/sites/prod/files/edg/media/RANDDINVESTMENTCRITERIA.pdf>
- OECD (2013). *Principles for the governance of regulators, public consultation draft*. OECD Publishing, Paris.
- OECD (2014). *Industry and technology Policies in Korea*, OECD Publishing. Retrieved 2015, from <http://dx.doi.org/10.1789/9789264213227-en>
- Padilla, P. (2015). *Policy learning through strategic intelligence*. PhD thesis in department of Science, Technology and Policy Studies in University of Twente, the Netherlands.
- Palumbo, D. J., & Hallett, M. A. (1993). Conflict versus consensus models in policy evaluation and implementation. *Evaluation and Programme Planning*, 16(1), 11–23.
- Peters, B., & van Nispen, F. (1998). *Pubic policy instruments evaluating the tools of public adminstration*. In Bemelmans–Videc, M., Rist R., & Vedung, E. (eds.): Carrots, sticks and sermons: policy instruments and their evaluations. New Jersey, USA: Transation Publishers.
- Poo, M., & Wang, L. (2014). on CAS reform and Pioneer Action Plan. *National Science Review*, 2(1), 1–12.
- Porter, M. (1996). What is strategy? *Harvard Business Review*. Novermber–December 1996, 1–21.
- Preskill, H., & Torres, R. T. (1999). Building capacity for organisational learning through evaluative inquiry. *Evaluation*, 5(1), 42–60.
- Price, A. W. (2008). *Contextuality in practical reason*, Oxford University Press.
- Raina, R. S. (2003). Disciplines, institutions and organisations: impact assessments in context. *Agricultural Systems*, 78(2), 185–211.
- Rebien, C. C. (1996). Participatory evaluation of development assistance: dealing with power and facilitative learning. *Evaluation*, 2(2), 151–171.
- Renders, A., & Gaeremynck, A. (2012). Corporate governance, principal–principal agency conflicts, and firm value in European listed companies. *Corporate Governance International Review*, 20(2), 125–143.
- Rich, R. F. (1997). Measuring knowledge utilisation: processes and outcomes. *Knowledge and Policy*, 10(3), 11–24
- Rip, A. (2000). Higher forms of nonsense. *European Review*, 8(4), 467–486.

- Röbbecke, M., & Simon, D. (2002). *The assessment of Leibniz institutes: the relationship between external and internal evaluation*. In Shapira, P. and Kuhlmann, S. (eds.): *Learning from science and technology policy evaluation: Proceedings from the 2000 U.S.-European Workshop*, Atlanta: School of Public Policy (Georgia Institute of Technology), and Karlsruhe: Fraunhofer Institute for Systems and Innovation Research, 290–297.
- Rodriguez, H., Fisher, E., & Schuurbiers, D. (2013). Integrating science and society in European Framework Programmes: Trends in project-level solicitations. *Research Policy*, 42(5), 1126–1137.
- Ross, J. A., Barkaoui, K., & Scott, G. (2007). Evaluations that consider the cost of educational programs—the contribution of high-quality studies. *American Journal of Evaluation*, 28(4), 477–492.
- Sanders, B. A. (2003). Maybe there's no such thing as a "good cop": organisational challenges in selecting quality officers. *Policing an International Journal of Police Strategies & Management*, 26(2), 313–328.
- Sanderson, I. (2002). Evaluation, policy learning and evidence-based policy making. *Public Administration*, 80(1), 1–22.
- Saunders, M. (2012). The use and usability of evaluation outputs: a social practice approach. *Evaluation*, 18(4), 421–436.
- Schiene, C. and Schimank, U. (2007), *Research evaluation as organisational development*, in Whitley and Gläser (eds.): *The changing governance of the sciences: the advent of research evaluation systems*, 171–188.
- Schimank, U. (2005). 'New public management' and the academic profession: reflections on the German situation. *Minerva*, 43(4), 361–376.
- Schimank, U. (1994). How German professors handled increasing scarcity of resources for their research: a three-level actor Constellation, in Uwe Schimank and Andreas Stuke (eds.): *Coping with trouble: how science reacts to political disturbances of research conditions*, New York: ST. Martin's Press, 35–59.
- Schruff, H. (2012). *Evaluation procedures of the Max Planck Society*. Retrieved 2013, from <http://www.cnr.it/sitocnr/UPO/documenti/ep2223ott.pdf>
- Science in Transition website (2015). Retrieved 2015, from <http://www.scienceintransition.nl/>
- Scriven, M. (1996). Types of evaluation and types of evaluator. *Evaluation Practice*, 17(2), 151–161.

- Shapira, P., & Kuhlmann, S. (2003). *Introduction: learning from science and technology policy evaluation*. In P. Shapira & S. Kuhlmann (eds.): *Learning from science and technology policy evaluation*. Kalsruhe, Germany, Atlanta, USA: Edward Elgar Publishing Limited, 1–15.
- Shove, E. (2003). Principals, agents and research programmes. *Science and Public Policy*, 30(5), 371–381.
- Simon, D., & Knie, A. (2013). Can evaluation contribute to the organisational development of academic institutions? an international comparison. *Evaluation*, 19(4), 402–418.
- Smith, A. J. (2004). Scientific freedom. *Journal of Dental Research*, 83(12), 895–895.
- Smits, R., and Kuhlmann, S. (2004). The rise of systemic instruments in innovation policy. *International Journal of Foresight and Innovation Policy*, 1(1–2), 4–32.
- Strassheim, H., & Kettunen, P. (2014). When does evidence-based policy turn into policy-based evidence? configurations, contexts and mechanisms. *Evidence & Policy*, 10(2), 259–277.
- Strong, B., Davenport, T., Prusak, L. (2008). Organisational governance of knowledge and learning, *knowledge and process management* 15(2), 150–157.
- Sun, Y. T., & Cao, C. (2014). Demystifying central government R&D spending in China. *Science*, 345(6200), 1006–1008.
- Suttmeier, R. P., Cao, C., & Simon, D. F. (2006). “Knowledge innovation” and the Chinese Academy of Sciences. *Science*, 312(5770), 58–59.
- Tampieri, L. (2013). The governance of synergies and conflicts in project management: The case of IPA project recoURB. *the Knowledge Economy*, 4(4), 370–386.
- Teece, D. J., Pisano, G., & Shuen, A. (1997). Dynamic capabilities and strategic management. *Strategic Management Journal*, 18(7), 509–533.
- Tsay, A., Lamont, M., Abbott, A., & Guetzkow, J. (2003). From character to intellect: changing conceptions of merit in the social sciences and humanities, 1951–1971. *Poetics*, 31(1), 23–49.
- Van der Knaap, P. (2006). Responsive evaluation and performance management: Overcoming the downsides of policy objectives and performance indicators. *Evaluation*, 12(3), 278–293.

- Van der Meulen, B. (1998). Science policies as principal–agent games—institutionalisation and path dependency in the relation between government and science. *Research Policy*, 27(4), 397–414.
- Van der Meulen, B., & Rip, A. (1998). Mediation in the Dutch science system. *Research Policy*, 27(8), 757–769.
- Van der Meulen, B., & Rip, A. (2000). Evaluation of societal quality of public sector research in the Netherlands. *Research Evaluation*, 9(1), 11–25.
- Van der Knaap, P. (2004). Theory-based evaluation and learning: possibilities and challenges. *Evaluation*, 10(1), 16–34.
- Van der Knaap, P. (2006). Policy evaluation and learning: feedback, enlightenment or argumentation? *Evaluation* 1 (2), 193–220.
- Walhout, B., & Kuhlmann, S. (2013). *In search of a governance framework for responsible research and innovation*. Paper presented at the IEEE International Technology Management Conference & 19th ICE Conference, The Hague, Netherlands.
- Way, M., & Ahmad, S. A. (2013). The San Francisco Declaration on Research Assessment. *Journal of Cell Science*, 126(9), 1903–1904.
- Whitley, R., & Gläser, J. (2007). *The changing governance of the sciences*: Springer.
- Wholey, J. (1983). *Evaluation and Effective Public Management*. Boston: Little, Brown.
- Wilsdon, J., et al. (2015). *The Metric tide: report of the independent review of the role of metrics in research assessment and management*. The Higher Education Funding Council for England. Retrieved September, 2015, from <http://www.hefce.ac.uk/pubs/rereports/Year/2015/metrictide/Title,104463,en.html>
- Wolf, H. G. (1995). An academy in transition—organisational success and failure in the process of German unification. *Social Studies of Science*, 25(4), 829–852.
- WR evaluation report (2001), *Statement of the German Science Council on the Hermann von Helmholtz Association of German Research Centres*, Köln, Germany. Retrieved 2014, from <http://www.wissenschaftsrat.de/download/archiv/4755-engl.pdf>
- Yanow, D. (1998). Studying museum buildings as organisational spaces while reflecting on interpretive methods and their narration. *Journal of Management Inquiry*, 7(3), 215–239.
- Yin, R. K. (2009). *Case study research design and methods*, US. SAGE Publications.

- Young, M. N., Peng, M. W., Ahlstrom, D., Bruton, G. D., & Jiang, Y. (2008). Corporate governance in emerging economies: a review of the principal–principal perspective. *Journal of Management Studies*, 45(1), 196–220.
- Zhang, D., Banker, R. D., Li, X., & Liu, W. (2011). Performance impact of research policy at the Chinese Academy of Sciences. *Research Policy*, 40(6), 875–885.
- Zhang, Q., Chen, L., & Feng, T. (2014). Mediation or moderation? the role of R&D investment in the relationship between corporate governance and firm performance: empirical evidence from the Chinese IT industry. *Corporate Governance and International Review*, 22(6), 501–517.
- Ziman, J. (2003). Non-instrumental roles of science. *Science and Engineering Ethics*, 9(1), 17–27.

Summary

Evaluation has been designed and used as an instrument of organisational governance for publicly funded research institutions (PRIs). Such an instrument can justify external public support and provide internal evidence for decision-making and organisational learning. Under given national and institutional contexts, defined missions of PRIs shape the characteristics of organisational governance in at least three aspects: strategies, funding, and operation. Governance tensions related to them involve key stakeholders at three organisational levels: central decision-making and supervisory bodies at the macro-level (L1), research institutes at the meso-level (L2) and individual researchers at the micro-level (L3).

The thesis offers a conceptual framework, built in a formative way, to answer the research question: How do evaluation mechanisms of PRIs play a balancing role and help to mitigate tensions in organisational governance? Three PRIs were investigated, namely, the Max Planck Society (MPG) and the Helmholtz Association (HGF) in Germany, and the Chinese Academy of Sciences (CAS). Their governance tensions across the three levels are reflected and balanced by the evaluation mechanisms, which facilitate the organisational learning. The balancing processes occur when the individual evaluation activities supplement each other within the institutionalisation of the overarching evaluation mechanisms. During the processes, the stakeholders' interactions and negotiations vis-à-vis their diverging interests are also facilitated.

Generalisable balancing roles of evaluation mechanisms in organisational governance are found at each organisational level in the three PRIs studied. Specifically, the balance found at L1 is between external justification of public funding and internal governance of heterogeneous research; at L2 between required alignment of organisational strategies and individual institutes' autonomous development; and at L3 between meeting management instructions and protecting scientists' trust and freedom. A common principle to mitigate governance tensions across the three levels involving various stakeholders is found—a proper contextualisation of evaluation designs and uses. Implications are suggested for useful ways of discovering and balancing governance tensions at PRIs in general and for public policies on research evaluation in wider landscapes of PRIs.

Samenvatting

Onderzoeksevaluatie wordt ontworpen en toegepast als governance instrument voor publiek gefinancierde onderzoeksinstituten (Publicly funded Research Institutes, PRIs). Als instrument kunnen evaluaties gebruikt worden om publieke ondersteuning te rechtvaardigen en data te leveren voor besluitvorming en organisationele leerprocessen. Binnen gegeven nationale en institutionele randvoorwaarden bepalen de missies van PRIs de governance karakteristieken van deze organisaties op ten minste drie aspecten: strategie, financiering en operaties. Bij spanningen op het gebied van governance zijn belanghebbenden op drie verschillende organisatieniveaus betrokken: centrale besluitvormings- en toezichtorganen op macroniveau (L1), onderzoeksinstituten op mesoniveau (L2) en individuele onderzoekers op microniveau (L3).

Dit proefschrift biedt een initieel conceptueel kader om de volgende onderzoeksvraag te beantwoorden: Hoe spelen evaluatiemechanismen van PRIs een balancerende rol en helpen zij om spanningen in organisationele governance te verlichten? Voor dit proefschrift zijn drie PRIs onderzocht: de Max Planck Society (MPG) en de Helmholtz Association (HGF) in Duitsland en de Chinese Academy of Sciences (CAS). Hun governance spanningen op de drie organisatieniveaus zijn weerspiegeld in, en worden gebalanceerd door, evaluatiemechanismen. Deze mechanismen ondersteunen organisationeel leren. Balanceringsprocessen vinden plaats wanneer de individuele evaluatieactiviteiten elkaar aanvullen binnen de institutionalisering van de overkoepelende evaluatiemechanismen. Tijdens deze processen worden de interacties en onderhandelingen van belanghebbenden en hun uiteenlopende perspectieven ook ondersteund.

Op ieder organisatieniveau van de drie bestudeerde PRIs zijn generaliseerbare rollen gevonden van evaluatiemechanismen ten aanzien van het balanceren van afwegingen en perspectieven in organisationele governance. De balans op niveau L1 ligt tussen externe rechtvaardiging van publieke financiering en interne governance van heterogeen onderzoek. De balans op niveau L2 ligt tussen de benodigde koppeling tussen organisatiestrategieën en de autonome ontwikkeling van individuele instituten. De balans op niveau L3 ligt tussen het tegemoetkomen van managementeisen en het beschermen van het vertrouwen en de vrijheid van individuele wetenschappers. Dit proefschrift vond een gemeenschappelijk principe om governance spanningen te verlichten op de drie niveaus waarbij diverse

belanghebbenden zijn betrokken, namelijk een juiste contextualisering van evaluatie ontwerpen en toepassingen. Het proefschrift schetst implicaties hiervan voor methodes waarbij governance spanningen bij PRIs in het algemeen kunnen worden ontdekt en worden afgewogen, en implicaties voor publiek beleid op het gebied van onderzoeksevaluaties in de bredere context van PRIs.