Department of Social Research
University of Helsinki
Finland

Interdisciplinary Accountability in the Evaluation of Research Proposals

Prospects for academic quality control across disciplinary boundaries

Katri Huutoniemi

ACADEMIC DISSERTATION

To be presented, with the permission of the Faculty of Social Sciences of the University of Helsinki, for public examination in Auditorium XIII, University main building,
on 30 November 2012, at 12 noon.

Helsinki 2012

Publications of the Department of Social Research 2012:17
Social and Public Policy

Cover: Jere Kasanen
Photo: Kirsi Nuppola / Juha Kiviluoma

Distribution and Sales:
Unigrafia Bookstore
http://kirjakauppa.unigrafia.fi/
books@unigrafia.fi
PL 4 (Vuorikatu 3 A) 00014 Helsingin yliopisto

ISSN-L 1798-9140
ISSN 1798-9132 (Online)
ISSN 1798-9140 (Print)
ISBN 978-952-10-7661-9 (Print)
ISBN 978-952-10-7662-6 (Online)

Unigrafia, Helsinki 2012

Abstract

This dissertation investigates academic research evaluation from the novel perspective of interdisciplinary accountability. While the standard model of evaluation puts a premium on disciplinary expertise and professional control, increasing demands for both interdisciplinarity and accountability have brought about pressures to open scholarly knowledge production to scrutiny beyond disciplinary boundaries. This study is concerned with the socio-epistemic implications of these developments, and discusses interdisciplinary accountability as an essential, yet underdeveloped mechanism of academic quality control. It asks what constitutes interdisciplinary accountability, and how it can be demonstrated, validated, and strengthened in the evaluation of research proposals.

The empirical part of the study focuses on the evaluation of research proposals in a national research funder in Finland, the Academy of Finland. Drawing on analyses of research proposals and peer review deliberations, the study explores the various ways in which scholars coordinate, negotiate, and modify different disciplinary regimes in the pursuit of high-quality scientific knowledge. Based on the empirical findings and a review of the literature on interdisciplinarity, social epistemology, and science policy, the study emphasizes the importance of considering epistemic accountabilities in a context-sensitive, open-ended manner in knowledge production and evaluation.

The study makes both a theoretical and a practical contribution. First, it provides a complementary perspective on the changing governance of science by articulating the notion of interdisciplinary accountability. While recent debates have emphasized problem solving and public accountability as important indicators of legitimate science today, this study argues that accountability across academic disciplines holds an equal promise of more relevant and reliable knowledge. Interdisciplinary accountability is thus a socio-epistemic mechanism for responsible science, and serves as a counterforce to disciplinary autonomy as well as the “tyranny” of political or economic forces over epistemic values.

Second, the study makes a practical contribution to the evaluation of interdisciplinary research. To this end, it articulates a framework for conceptualizing interdisciplinary accountability in research proposals, and considers ways to include interdisciplinary accountability in peer review. The framework helps to identify the relevant epistemic stakeholders, the functions and benefits of proposed research, as well as the methodological procedures for accomplishing the stated goals, which constitute the prerequisite for any evaluative act. As for the evaluative act itself, the study suggests using interdisciplinary dialog between reviewers as a type of epistemic standard. A reasonable strategy is to mix experts from different but not disparate fields, and select generalist panel members who possess a broad knowledge beyond any one academic field.

Tiivistelmä

Tieteellistä laatua on pääsääntöisesti arvioitu kunkin tieteenalan omista lähtökohdista käsin, mutta tieteidenvälisen yhteistyön ja tieteen yhteiskunnallisen vastuuvelvollisuuden vaatimukset ovat luoneet paineita tieteenalarajat ylittävälle tiedontuotannon hallinnoinnille. Tutkimus tarkastelee tieteellisen arvioinnin tavoitteita ja käytäntöjä tästä näkökulmasta, ja tuo keskusteluun tieteidenvälisen vastuuvelvollisuuden käsitteen. Tutkimuksessa kysytään, mitä tieteidenvälinen vastuuvelvollisuus pitää sisällään, ja miten se voidaan osoittaa, todentaa ja ottaa huomioon tutkimushankkeiden arvioinnissa.

Tieteidenvälisen vastuuvelvollisuuden ilmenemistä tarkastellaan Suomen Akatemian tutkimushankkeiden arvioinnissa. Empiirinen tutkimus kohdistuu tieteidenvälisen vuorovaikutuksen muotoihin yhtäältä hankesuunnitelmien sisällössä ja toisaalta niiden vertaisarviointiprosessissa. Analyysien kohteena on se, miten tutkijat ja arvioitsijat aktiivisesti koordinoivat, sovittelevat ja muokkaavat tieteenalojen asettamia normeja pyrkiessään tieteellisesti korkeatasoiseen tutkimukseen. Empiiristen löydösten sekä tieteidenvälisyyttä, tieteentutkimusta ja tiedepolitiikkaa käsittelevän kirjallisuuden perusteella esitetään, että episteemisten vastuuvelvollisuuksien tapauskohtainen harkinta on keskeinen elementti uuden tiedon tuotannossa ja arvioinnissa.

Tutkimus tuottaa täydentävän näkökulman tiedontuotannon hallinnointia ja sen muutoksia koskevaan keskusteluun, jossa on viime aikoina korostunut tieteen yhteiskunnallinen vastuuvelvollisuus ja tieteellisen tiedon hyödynnettävyys. Työssä esitetään, että tieteidenvälisten suhteiden arvioiminen on ensijainen, mutta vähälle huomiolle jäänyt osa vastuullista tiedontuotantoa. Tieteidenvälinen vastuuvelvollisuus asettaa tieteenaloittaisen tutkimuksen laajemman tiedeyhteisön arvioitavaksi, ja pyrkii siten parantamaan tutkimustiedon luotettavuutta ja tieteellistä relevanssia. Työn keskeiset tulokset tukevat tämän näkökohdan operationalisointia ja edistämistä tutkimushankkeiden arvioinnissa.

Arvioinnin viitekehykseksi tutkimus tarjoaa käsitteellisen jäsennyksen tieteidenvälisen vastuuvelvollisuuden rakenteesta. Jäsennys auttaa määrittämään yksittäisen tutkimushankkeen tieteidenvälisiä arviointiperusteita kolmella ulottuvuudella: mille tieteenaloille, minkälaisesta tutkimustavoitteesta, ja minkälaisesta tutkimusprosessista hankkeessa ollaan vastuuvelvollisia. Arvioinnin toteutuksessa tutkimus korostaa tieteidenvälisen neuvotteluympäristön rakentamista. Tieteellisten asiantuntijoiden valinnalla ja asiantuntijapaneelin tieteenalakokoonpanolla voidaan ohjata sitä, missä määrin arvioitsijat rakentavat toistensa asiantuntemuksen ja arvostusten varaan muodostaessaan käsitystä tutkimushankkeiden laadusta.

Preface

This work, like most others, can best be understood against the intellectual background of the author. My view of research in any field of science is that of an observer; I never felt that I am an expert in social science, for example, as I entered the Department of Social Research at a late stage of my research process. Similarly, while I did my master's degree in Environmental Science and Policy, I focused more on the epistemological basis of the field than on the subject matter itself. One might think that perhaps I am a philosopher or epistemologist by heart, but that is not quite true either. I simply do not see a reason to fully acquaint myself with any disciplinary practice, including the philosophy of science. At the same time, several organizational and faculty changes during my doctoral studies made it difficult for me to establish a firm relationship with any university department.

Due to this background, I have had a continuous struggle with the academic relevance of my research. It has been hard to identify key professional networks or target audiences, which has given me a weird feeling of not taking responsibility of my work. The relief was double when I was able to articulate the notion of interdisciplinary accountability as the key concept of this dissertation. Besides providing a theoretical perspective that links my individual articles together, it also reflects my deepest intellectual and moral stance as a researcher. Such a stance would probably not occur—and may be more difficult to make sense—to those who perceive themselves as professionals. I hope, however, that my ideas of interdisciplinary accountability would spoke to those readers as well, and invoke their concern for the overall goals of science.

This dissertation would not have been possible without several organizations and indiduals. During the seven-year period (which includes two maternity leaves), one of the few steady things was funding from the Finnish postgraduate School in Science, Technology and Innovation Studies (TITEKO). Other funders include the Academy of Finland (the Research Council for Culture and Society offered a grant for research training abroad for one academic year) and Emil Aaltonen Foundation. Another steady support was my supervisor Janne Hukkinen, who not only gave me invaluable advice on my work but also an exemplary model of an interdisciplinary thinker. Janne was also leading the two weekly doctoral seminars that most influenced the scholarly frames of this research.

Besides Janne, I had the privilege to work with three other wonderful supervisors: Henrik Bruun, Julie Thompson Klein, and Michèle Lamont. Henke was so devoted, inspiring, and close to my research interests that I could not resist his suggestion to start a PhD, and despite being disappointed by his departure from the academia, I owe him a lot of gratitude of the path I chose. Julie has given me enduring support mainly from a distance—introduced me for important persons, written a number of recommendation letters, helped me to link with relevant literatures, etc. Her broad view of interdisciplinarity also inspired my research from the very start. One of the key persons whom I would not have met without Julie is Michèle, who kindly hosted me during my stay at Harvard University. I am truly indebted and thankful to Michèle for sharing her brilliant ideas and observations about academic judgment and the sociology of knowledge and evaluation. I am also thankful for her overall kindness and care of me during what turned out to be quite difficult year.

I would like to thank the Academy of Finland not only for awarding me a research grant, but also for initiating my empirical research on their materials and procedures in the first place, and permitting me to continue and broaden the research beyond their own organizational interests. While my personal access to the meetings of evaluation panels was ultimately not allowed, the personnel were very supportive of my goal to investigate the confidential procedures at close proximity. I am grateful to Annamaija Lehvo and Riitta Mustonen who helped with many practical issues; Tiina Forsman, Mirka Gustafsson, Heli Karjalainen, Kustaa Multamäki, and Jaana Vormisto for helping to contact expert panelists; and Paavo Löppönen, Anneli Pauli, and Meri Vannas for sorting out my research permissions. In addition, I want to thank the expert panelists who shared their experiences of the peer review process, and the funding officers who explained me the evaluation prodecure used by the Academy.

Colleagues at several universities and departments deserve warm thanks for support and feedback. The most recent group of colleagues include those who attended the Environmental Policy Research Seminar at the Department of Social Research at the University of Helsinki: Annukka Berg, Eeva Berglund, Nina Honkela, Jarkko Levänen, Paula Saikkonen, Arho Toikka, Johan Munck af Rosenschöld, Sami Heikkilä, Antto Vihma, Sarianne Tikkanen, and others. Other people on the third floor of Snellmaninkatu 10 were nice company as well. Before I entered the environmental policy research group, I had a privilege to work with most inspiring people in the Laboratory of Environmental Protection at the Helsinki University of Technology—in addition to Janne, Henke, and Nina, I would like to thank Richard Langlais, Mikko Rask, Olli Salmi, Aino Toppinen, Maria Höyssä, Martti Timonen, Anu Tuominen, and others. Yet another important forum of debate has been the regular seminars and summer schools of TITEKO, which provided a good opportunity to engage with the field of Science and Technology Studies.

The intellectual roots of this dissertation are, in part, in the subject of Environmental Science and Policy at the Department of Environmental Sciences at the University of Helsinki. Its open and democratic learning tradition encouraged me to follow quite non-disciplinary line of study, and I want to thank especially the legacy of “KVYST” (“kokonaisvaltainen ympäristönsuojelutiede” in Finnish) for providing me some initial ideas of this work. After those years, occasional feedback and support from Petri Tapio and Riikka Paloniemi have been helpful. I am also grateful for the recent reunion with many KVYST people around a collaborative book project that started during the final stages of this dissertation.

I owe sincere gratitude to the pre-examiners Terttu Luukkonen and Robert Frodeman, whose concise and insightful comments gave this dissertation its final form and title. Finally, I am much obliged to my wonderful friends and family, not least my little sons, who have made my life a life, not research. To my husband Juha I owe so much for believing in me, even when I did not.

Helsinki, 30 October 2012
Katri Huutoniemi

Contents

List of original publications

This thesis is based on the following publications:

I Huutoniemi K (2010): Evaluating Interdisciplinary Research. In: Frodeman R, Klein JT and Mitcham C (Eds), *The Oxford Handbook of Interdisciplinarity.* Oxford: Oxford University Press. Pp. 309–320.

II Huutoniemi K, Klein JT, Bruun H and Hukkinen J (2010): Analyzing Interdisciplinarity: Typology and Indicators. *Research Policy* 39(1): 79–88.

III Lamont M and Huutoniemi K (2011): Comparing Customary Rules of Fairness: Evaluative Practices in Various Types of Peer Review Panels. In: Camic C, Gross N and Lamont M (Eds), *Social Knowledge in the Making*. Chicago, IL: University of Chicago Press. Pp. 209–232.

IV Huutoniemi K (2012): Communicating and Compromising on Disciplinary Expertise in the Peer Review of Research Proposals. *Social Studies of Science* 42(6): 900–924.

The publications are referred to in the text by their roman numerals.

Contributions of the author

Article I and *Article IV* are solely the work of the author. The author had the main responsibility for authoring *Article II* as a whole, but it draws on a collaborative project. In that project, the author contributed in designing the work and developing the conceptual framework, and was responsible for conducting the empirical analysis. In *Article III*, the author was responsible for one of two empirical studies that were paralleled in the paper. The author also participated in designing the comparative analysis and drawing the conclusions.

1 Introduction

As a master's student of the interdisciplinary subject of environmental science and policy, I learned to question the tunnel vision of many academic disciplines and focus instead on the complex nature of environmental problems. Later on I realized that many programmatic claims about interdisciplinary science offered a tunnel vision as well, but of another kind—what is not open to scrutiny from outside is not accountable. These observations made me think about epistemic accountabilities across and within disciplines more generally, and inspired me to explore how scholars actually account for their interdisciplinary practices in the pursuit of new knowledge. Drawing on the expanding literature on interdisciplinarity as well as my own empirical studies of the evaluation processes in a national research funder in Finland, this dissertation considers pragmatic means for opening scholarly knowledge production to scrutiny beyond disciplinary boundaries. It discusses interdisciplinary accountability as an essential yet underdeveloped element in the governance of science,[1] and suggests concepts and practices for incorporating more interdisciplinary approaches in the policies of research evaluation.

While the classical approach to science considered the pursuit of truth to be the final goal of science, this search has been gradually replaced by the more pragmatic goal of producing reliable and relevant knowledge. The constitution of such knowledge in the era of knowledge society is, however, more heterogeneous and ambiguous than ever before (Nowotny et al. 2002). This means not only that multiple and often inconsistent perspectives on the same issues and concerns may be equally sound, but also that we are increasingly pressed to take action on the basis of such dissonant knowledge. Many of the pressing problems of contemporary society do not allow for the peaceful coexistence of incommensurable views, but call for coordinated solutions, however temporary and partial. Under the current volatility of natural environment, rapid technological change, and increasing complexity of our societies, a major intellectual challenge of academia is to cope with the dissonance of knowledge.

Recent debates on the knowledge society have emphasized the contextualization of problems and public accountability as important indicators of knowledge robustness (e.g. Gibbons & Nowotny 2001; Nowotny et al. 2002; Nowotny 2003; Maasen & Lieven 2006; Maasen et al. 2006). According to the current understanding of the concept, "accountability" refers to a demonstration that science has taken society into account, and that society is not simply the recipient of the knowledge, but has input into the science (Strathern 2004). At the same time, there is little discussion about accountability across disciplines. Interdisciplinarity is sometimes portrayed as an index of accountability in knowledge production, as "it is an implicit evaluation of the success of disciplines to convey their messages" (Strathern 2004, 79). This notion, however, builds on the above idea of public accountability, of which interdisciplinarity is only a marker. The goal of this dissertation is to make explicit what has been left unexplored by Marilyn Strathern and

[1] *By "science" I refer, throughout this dissertation, to the systematic pursuit of knowledge, which includes all of the academic disciplines, not just the natural sciences.*

others: What constitutes interdisciplinary accountability, and how can it be demonstrated, validated, and strengthened?

The starting point of this study is that in addition to—or preceding—the current demands for public accountability in knowledge production is the demand for interdisciplinary accountability. While the major goal of public accountability is the *usability* of knowledge, I understand interdisciplinary accountability primarily as an *epistemic* mechanism—what it promises is more robust, reliable, and relevant knowledge.[2] Thus, this work is concerned with rendering academic disciplines more accountable to each other in order to sustain the power of scientific knowledge to persuade. This includes the necessity to open disciplinary knowledge production—not only its deliverables, but essentially its goals and procedures—to scrutiny from all quarters. Interdisciplinary accountability, as discussed in this dissertation, is understood as the willingness of and means available to researchers to be more responsive to the scientific community at large, not only to their disciplinary colleagues. This stance comes close to understanding interdisciplinarity as a philosophy of knowledge (Frodeman 2010a & 2011; Morin 2008) that challenges the disciplinary mode of producing and evaluating knowledge. It does not mean rejecting rigor or abandoning standards, but broadening the intellectual context in which they are defined.

Accountability refers to the process of "giving an account" or being answerable or capable of being accounted for (Alkin 2004). It is at once a moral stance towards the wider world and a set of procedures for verification (Strathern 2004). The concept has a long tradition in political science and finance, but its central idea is more general: "When decision-making power is transferred from a principal (e.g. the citizens) to an agent (e.g. government), there must be a mechanism in place for holding the agent to account for their decisions and if necessary for imposing sanctions, ultimately by removing the agent from power" (Lindberg 1999, 1). This dissertation analyzes interdisciplinary accountability as a necessary, but insuffiently developed element of the governance of scientific knowledge production. In particular, I ask how interdisciplinary accountability manifests itself, and can be strengthened, in the *evaluation of research proposals*.

Research evaluation is an essential means of exercising control over science. It is based on account-giving mechanisms, which actively shape assumptions about accountability and are dependent on accounting practices. Research evaluation in the modern sciences has traditionally rested on the idea of disciplinary quality control, the purpose of which has been to certify research activity as valid and reward those researchers who have produced knowledge. In the context of an increasingly present "audit culture" in higher education and research (Strathern 2000a) as well as the "audit society" more generally (Power 1997), the internal accountability of disciplines has been accompanied by a greater expectation of public accountability. Underlying this change is

[2] *In this dissertation, I use the terms "accountability", "accountable", etc. mainly in this epistemic sense. Obviously, other forms of accountability are also relevant in scientific knowledge production and evaluation. These include, but are not restricted to, legal accountability, which concerns the various laws on the ethical conduct of science, and admistrative accountability, which concerns the formal relationships between actors or activities at different organizational levels.*

the regime of "New Public Management", which has involved a reduction of hierarchical control in public sector organizations in favor of more decentralized governance based on setting broad objectives and then relying on *ex post* monitoring and evaluation of performance (Braun & Merrien 1999; Ferlie et al. 2009). The evaluation of research proposals in many funding organizations, for example, has become a forum of policy debate, with increasing emphasis on the strategic importance and socio-economic impacts of research activities (Luukkonen 2002).

In the face of these new accountabilities in knowledge production, research evaluation as an *epistemic*[3] mechanism of governance is at risk of losing its power. It is sometimes claimed that science has been put merely to the service of political agendas, with the resulting risk of destroying the scientific enterprise in the long run (e.g. Ziman 1996). At the same time, it seems obvious that the ideal of an autonomous discipline has come to its end; due to heightened specialization, science is losing sight of overall goals (e.g. Frodeman 2010a). The evaluation of interdisciplinary research is a case in point. While interdisciplinarity is highly prized by policymakers and research funders, recent history and numerous policy debates show that interdisciplinary research evaluations are often marked by conflict over what interdisciplinarity is, what criteria should be used to evaluate it, and what constitutes a legitimate evaluation procedure (e.g. National Academy of Sciences 2005; *Article I*). Below are some examples that invite us to consider epistemic accountabilities in science.

First, whenever research crosses boundaries between disciplines, the problem arises that each discipline carries specific and sometimes conflicting assumptions about quality. The criteria of disciplinary communities are proving insufficient for research that expands, integrates, or challenges the discipline's own canon. As Steve Fuller (2002) has asked: How does one judge the relative merit of *importing* ideas and findings from another discipline into one's own compared to the merits of *exporting* ideas and findings from one's own discipline into another? Or, what value is placed on work that is explicitly critical in intent, such as the reanalysis of data, the replication of an experiment, or a methodological or theoretical "audit" of a field? In such intellectual exchanges, what exactly is it that decides the matter of relevance: one's own discipline or the other discipline—or some combination of the two? Uncertainties of this kind leave interdisciplinary research with an unsettled epistemic status. On the one hand, many scholars who struggle to uphold stringent academic standards think that interdisciplinary research cannot be trusted to produce reliable knowledge; it is often denounced as being of dubious quality (Boix Mansilla 2006; Weingart 2000). On the other hand, interdisciplinary aspects of research may be placed outside of any scrutiny, as exceptional or meritoriously

[3] *While emphasizing this aspect, I do not think that there are "pure" epistemic categories. On the contrary, I wish to expand epistemology's horizons to various social, cognitive, political, and ethical dimensions of justified knowledge. Throughout the text, I use the terms "epistemic" and "epistemological" in this broader sense. "Epistemic" refers to matters pertaining to justified knowledge, whereas "epistemology" refers to a particular view (e.g. that of a discipline) or study (e.g. social epistemology) of epistemic questions.*

exempt. Both tendencies are obviously problematic for the epistemic functions of research evaluation.

Second, and related to the previous point, it is unclear *who judges* interdisciplinary work. Since there is no clearly defined community of peer reviewers as there often is in disciplinary[4] quality control, qualified reviewers can be very hard to find. With a small pool of potential reviewers, the difficulties of matching technical specialties while avoiding conflicts of interest increases dramatically (e.g. Eisenhart 2002). Beyond such pragmatic constraints, peer review is often deemed biased towards established approaches (e.g. Chubin & Hackett 1990; Langfeldt 2004; Porter & Rossini 1985), unreliable in assessing interdisciplinary research (e.g. Travis & Collins 1991), or relatively useless in helping one to make choices between different research fields (Fuller 2002; Weinberg 1962). As peer review is, however, the major mechanism of validating knowledge and distributing resources in academia—such as research grants, scholarships, jobs, journal space, etc.—the above problems clearly undermine its effective and equitable functioning. Especially in the face of decreasing budgets for research funding, it has become necessary to choose between competing types of high-quality research and compare different research areas against each other.

Third, there is no consensus on what constitutes interdisciplinarity, and how it can be identified in practice (*Article II*; Bruun, Hukkinen et al. 2005). Despite decades-long scholarly work on the concept of interdisciplinarity, no general interdisciplinary indicator useful for the purposes of research evaluation has been accepted (Porter et al. 2006). For example, in its first call for proposals for the *Training and Mobility of Researchers* program, the European Commission (EC) established a panel of evaluators from different backgrounds with the aim that they should assess proposals in terms of their interdisciplinarity, but no single project passed the review process due to the different evaluations given by members of the panel. The EC eliminated the interdisciplinary panel the following year (Rogers et al. 2005). Even researchers themselves seem unclear about whether their own work is interdisciplinary or not. While many researchers find themselves crossing the boundaries of disciplines, fields, or university departments in their work, they are not particularly comfortable calling what they do "interdisciplinary" or even "multidisciplinary" (Lattuca 2001; Palmer 2001). The definitional debate tends to be paralyzed by the notion that interdisciplinary research can have so many profiles (see Klein 2006 & 2008a).

As a response to these problems, a specific discourse devoted to the evaluation and criteria of interdisciplinary research has emerged (see *Article I*). However, it can be asked to what extent the problems that accompany interdisciplinary research are aberrations within an otherwise appropriate system of scientific knowledge production, or whether the problems are routine and symptomatic of much deeper challenges (see e.g. Frodeman 2011; Schwandt 2002). As we will see in the following sections, there are good arguments for the latter position: I suggest that these and other similar problems derive from a deficit of interdisciplinary accountability in the coordination and evaluation of knowledge

[4] *I use the term "disciplinary" as a counterconcept to "interdisciplinary"; it covers the meaning of "single disciplinary", "intra-disciplinary", and "unidisciplinary".*

production. The very meaning of interdisciplinarity, therefore, cannot be understood without challenging the authority of disciplinary norms.

My dissertation focuses particularly on interdisciplinary accountability in the evaluation of research proposals. My contribution is based on a review of the literature on the epistemic characteristics of interdisciplinarity, as well as science and technology studies and science policy literature. In addition, two rounds of empirical studies were conducted in the context of a public research funder in Finland, the Academy of Finland. The results of these studies are reported in four original articles, each of which offers specific insights on interdisciplinary accountability in research evaluation. The content of the articles are summarized in Table 1. Their specific contributions to this dissertation can be articulated as follows:

Article I opens up the problematics of this dissertation. By drawing on existing literature on the topic, it introduces the central challenges involved in evaluating interdisciplinary research. It focuses attention not on the criteria used to conduct interdisciplinary research, but on the perspectives used to evaluate it, and highlights the consequential role of both concepts and practices in defining merit: First, it shows how different conceptualizations of interdisciplinarity shape assumptions about quality; and second, it discusses how values are actively constructed by the people and practices involved. The article also correctly anticipated the position upheld in this dissertation, namely, that the evaluative issues raised by interdisciplinarity may indicate deeper challenges to traditional research evaluation, including limits in our current notions of accountability. Section 2.4, in particular, further elaborates the issues covered by this article.

Article II works towards an adequate conceptualization of interdisciplinarity as an object of analysis and evaluation. It addresses the question of how to define and identify, and thereby render accountable, interdisciplinary aspects of research without reducing the complexity and multiple meanings of the concept. To this end, it demonstrates a typology and indicators for analyzing interdisciplinarity in research proposals. The typology focuses on the intellectual or cognitive aspects of research rather than on formal institutions, and delineates interdisciplinarity as a routine part of scientific inquiry, instead of a category of its own. It thus sets the stage—but does not argue explicitly—for thinking about interdisciplinarity as a horizontal, constitutive part of accountability in knowledge production, one that challenges the disciplinary organization of knowledge. Section 4.1 discusses the findings of this article.

Article III is concerned with the social conditions conducive to achieving consensus about the quality of research proposals. It discusses informal practices that lead peer review panelists to regard their collective judgments as fair and legitimate and to a belief that they are able to identify optimal from suboptimal proposals. Thus, the article offers an analytical perspective for considering peer evaluations as occurring in a particular context of accountability and unfolding through the evaluators' interactions. Some of the practices discussed have a significant influence on whether and how reviewers of different disciplines held each other accountable for their own criteria and built their own evaluations on each others' judgments. Section 4.2, in particular, discusses the findings of this article.

Article IV builds on the same analytical approach as *Article III* but its specific goal is to understand and enhance interdisciplinary accountability in evaluation panels. The article compares the deliberative processes of different evaluation panels, and suggests that the disciplinary composition of a panel creates a particular sphere of socio-epistemological control and reciprocal accountability, which is a special case of the situationally shaped behavior observed in *Article III*. On this basis, the paper makes policy recommendations for strengthening interdisciplinary accountability in and by peer review. The findings of this paper are discussed especially in Section 4.3.

This dissertation summarizes the contributions of the four articles from the overarching perspective of interdisciplinary accountability in the evaluation of research proposals. The argument developed here builds on, but is different from, those made in the original articles. While two of the articles contribute to a specific discourse devoted to the evaluation and indicators of interdisciplinary research (*Articles I-II*), and the two others are primarily socio-cultural analyses of peer review (*Articles III-IV*), this dissertation takes a somewhat different position. It discusses interdisciplinary accountability as an essential, yet undervalued and underdeveloped element of the governance of scientific knowledge production. The overarching research question of this dissertation is: *What constitutes interdisciplinary accountability, and how can it be demonstrated, validated, and strengthened in the evaluation of research proposals?* I will answer this question by considering two sub-questions: (1) How to conceptualize interdisciplinary accountability in research proposals? (2) How can peer review facilitate interdisciplinary accountability?

The remainder or this dissertation is structured as follows. In Section 2, I present the theoretical underpinnings of the dissertation. I first consider the role of academic disciplines in the governance of scientific knowledge production and the way in which the increasing demands of both interdisciplinarity and accountability have challenged this situation. I then argue for interdisciplinary accountability as an epistemic mechanism of governance. From this standpoint, I summarize the contributions and shortcomings of the existing debate on evaluating interdisciplinary research, and define the problem space for the dissertation. In Section 3, I describe how I designed the study to answer the above research questions, and what data and methods I employed to fulfill this goal. While the specific methods applied in the original studies are described in the articles, the methodology section here only aggregates the data and explains how the synthesis was created. In Section 4, I respond to the two research questions on the basis of the original articles. I then discuss the findings in a more synthetic manner in Section 5, and consider the meaning of interdisciplinary accountability beyond the particular context of this study. In Section 6, I conclude by considering the implications for how research evaluations are conducted as well as possibilities for future research.

Table 1. *Summaries of the four dissertation articles.*

I Evaluating Interdisciplinary Research

The article analyzes the key characteristics and challenges of interdisciplinary assessment by drawing insights from the conceptual and pragmatic discussions of interdisciplinary research, empirical analyses of evaluation activities, and initiatives and experiences of participating organizations. It articulates three evaluative approaches to interdisciplinary research: (1) mastering multiple disciplines, (2) emphasizing integration and synergy, and (3) critiquing disciplinarity. It argues that these competing positions on interdisciplinarity shape assumptions about quality and how it should be evaluated, while the actual process of evaluation with various social, cognitive, and pragmatic aspects also plays an important role in quality judgments. The question is raised of whether and how the challenges of interdisciplinary assessment are distinct from the more general problematic of research evaluation today.

II Analyzing Interdisciplinarity: Typology and Indicators

The article presents a new typology and qualitative indicators for analyzing interdisciplinarity in research proposals. The proposed conceptual framework responds to the need for a robust and nuanced approach that is grounded in a deeper understanding of knowledge integration. As an example of using the framework, I and my co-authors discuss our classification of research proposals funded by the Academy of Finland. Our experience of using the framework also illustrates some interesting findings about interdisciplinarity. We found, for example, that the integrative pattern of interaction was more common than the multidisciplinary pattern; that a majority of interdisciplinarity was epistemically oriented rather than instrumentally oriented; and that a considerable amount of research was interdisciplinary to some extent.

III Comparing Customary Rules of Fairness: Evaluative Practices in Various Types of Peer Review Panels

The article analyzes and compares the intersubjective understandings that academic experts create and maintain in making collective judgments on research quality. The analysis is based on two parallel, but interconnected empirical studies, conducted in the United States and in Finland. The American study analyzed multidisciplinary funding panels in the social sciences and the humanities, and documented the customary rules that panelists use. The study of Finnish panels, in turn, compared a few evaluation panels in the environmental and social sciences, and examined how the composition of panels can influence customary rules. The dialogue between the studies points to some similarities and differences in the internal dynamics of peer review panels and sheds light on how evaluative settings enable and constrain evaluative behavior.

IV Communicating and Compromising on Disciplinary Expertise in the Peer Review of Research Proposals

The article compares peer review deliberations in four evaluation panels that differ in terms of scope and disciplinary heterogeneity. Based on evaluation reports and discussions with panel members, it illustrates a variety of ways in which reviewers bridge their different areas of expertise and achieve consensus on the quality of research proposals. The analysis demonstrates that peer review panels may be places where communication across disciplines occurs and interdisciplinary judgments arise, while disciplinary gatekeeping and incommensurabilities may impose limits on such communication. The comparison of deliberative processes sheds light on how collective judgments are shaped and constrained by the disciplinary design of the panel and by the intersubjective dynamics that unfold in deliberation. Based on these findings, the article considers conditions that may enhance disciplinary interaction as well as complementary judgments in proposal peer review, and thereby the prospects for interdisciplinary research.

2 Theoretical background

2.1 Disciplines and interdisciplinarity in scientific knowledge production

One of the main features of modern science is that it is sharply differentiated in terms of intellectual fields, also called disciplines—such as physics, chemistry, biology, psychology, sociology, economics, and so forth. Classifications of knowledge, of which modern academic disciplines are but one example, structure our perception of the world. They are deemed necessary for intellectual development, as they help prevent knowledge from becoming too abstract or overwhelming (e.g. Abbott 2001). Until the end of eighteenth century, disciplines served mainly as repositories of accepted knowledge, with their own respective subject matters, and functioned relatively independently. The institutionalization of academic disciplines, which involved the establishment of the current structure of universities, was thus a logical outcome of the functional specialization of science (Weingart 2010; Ziman 2000).

After their institutionalization, however, disciplines started to develop social functions of their own. The "natural monopoly" the pre-modern disciplines had on expertise in their respective subjects, and their autonomy in determining their own development (Weingart 2010), have turned into internally-driven specialization. That is, beyond the intellectual benefits of having a disciplinary organization of science, there have been several social and psychological mechanisms that maintain the division of labor between disciplines and strengthen the patterns of specialization (Campbell 1969; Gieryn 1999; Shadish & Fuller 1994; Turner 2000; Ziman 1997). An inevitable consequence of this specialization is a pluralism of epistemic cultures. Social scientific examination of the way actors go about producing knowledge reveals that each discipline operates with its own machinery of knowledge, constituting its own norms of production and evaluation (e.g. Knorr Cetina 1999; Lamont 2009).

This development has gone so far, it is argued, that the pursuit of specialization today lacks epistemic warrant: the upshot is that disciplines gain robust results, but within a self-contained bubble (Frodeman 2010a). As it has become permissible to restrict one's learning and expertise to a very narrow area (e.g. Abbott 2001), researchers have become unwilling or incapable of communicating beyond their own specialties. As a result of this self-perpetuating cycle, disciplinary knowledges appear more or less incommensurable with each other. The difficulties in integrating knowledge from various traditions of environmental research illustrate this discontinuity (Huutoniemi 2004). A firmly institutionalized example of the phenomenon is the neo-classical paradigm of economics, whose basic assumptions about the behavior of the individual human being psychologists have long ago shown to be incorrect (Fuller 1988, 194).

While this tendency has been somewhat unavoidable, there are also cognitive and political stakes involved in claims about incommensurability between disciplines. One of them is what Alberto Cambrosio and Peter Keating (1983) have called the "disciplinary stake". The disciplinary stake would be characterized by the power, held by the producers,

to define the doctrinal corpus to be transmitted, the rules of apprenticeship, and the methods of certification and sanction. The disciplinary stake is, therefore, to release the given scholarly practice from domination by competing disciplines, either "above" or "beside" it, and to dictate its own rules. What is ultimately at stake is influence, i.e. the power to define a given field of investigation over the definitions proposed by competing disciplines, since influence can translate into career opportunities and other resources (Turner 2000; Weingart 2010). At the same time, the claims, activities, and institutional structures that define and protect disciplinary practices implicitly undermine attempts to evaluate or compare the goals and achievements of different disciplines according to any common or external metrics (see Espeland & Stevens 1998). The institutional frontiers between disciplines at any given moment can thus be understood as the consequence of the reification of the "disciplinary stake".

The relations between disciplines, and the role of interdisciplinarity, are old topics in the sociology, philosophy, and history of science. In their critical review of interdisciplinarity, Jerry Jacobs and Scott Frickel (2009) conclude that disciplines and interdisciplinarity are not distinct systems, but neither is the relation between the two organizationally uniform or historically stable. They note, however, that the established disciplines are not as static or as isolated as advocates of interdisciplinarity sometimes suggest; they remain dynamic centers of knowledge production that are open to external developments even while insisting on internal standards (see also Bruun, Hukkinen et al. 2005). According to the authors, empirical evidence indicates that the traditional disciplines are not as dominant as they once were; they represent a smaller share of the academy than was the case only a generation ago. Following Richard Whitley's observation that "traditional patterns of integration and control through academic disciplines seem to have broken down in many fields without any coherent and stable structure emerging to replace them" (Whitley 1984, 292), they set increased interdisciplinarity directly against the persistence of disciplinary control.

The proliferation of interdisciplinary funding programs, institutes, and other science policy incentives to combine disciplinary resources indicates a clear counter-trend to the increasing specialization of science (e.g. Bruun, Hukkinen et al. 2005; Cunningham 1997; European Union Research Advisory Board 2004; National Academy of Sciences 2005). Underlying the current push for interdisciplinarity is the need to integrate knowledge and solve problems that individual disciplines cannot solve alone (e.g. Frodeman & Mitcham 2007; Klein 1990 & 1996). The inability of disciplinary knowledge to tackle many contemporary problems, in areas such as globalization, environment, health, and security, is well known, and interdisciplinarity is expected to offer more comprehensive solutions. A growing emphasis has also been put on the potential of interdisciplinary integration to foster scientific progress and creativity as well as economic and technological innovation (e.g. Bruce et al. 2004; Bruun & Toppinen 2004; National Academy of Sciences 2005; Stefik & Stefik 2004). The growth of interdisciplinarity is linked to another major trend in science—the increasing demand for accountability—which is discussed next.

2.2 Interdisciplinarity and the demands of accountability

After World War II, public funding of science grew tremendously across the Western world. Until fairly recently, this funding came with few strings attached. The assumption was that there was an automatic relation between scientific knowledge production and the social good (e.g. Pielke & Bylerly 1998). Optimally, it was argued, science would follow a set of institutional imperatives or norms that Robert K. Merton ([1942] 1973) identified as communalism, universalism, disinterestedness, and organized skepticism. Largely due to this ethos, science was thought to progress best when scientists are left alone to pursue the questions that interest them—as Michael Polanyi famously argued in *The Republic of Science* (1962).

Over the past few decades, major changes in the governance of higher education and research have taken place in many OECD countries. National budgets for research funding throughout the Western world have reached their "limits to growth", and science has undergone a radical structural transformation to a much more tightly organized, rationalized, and managed social organization (Ziman 1994). These changes have altered the nature of the power relationships governing research priorities and the evaluation of results (Braun & Merrien 1999; Ferlie et al. 2009). In particular, increasingly exogeneous and formalized nature of the mechanisms of governance, as well as the strength and extent of their enforcement (Whitley 2011), have broken the autonomy of disciplinary communities and of academic organizations with regard to their goals and procedures. The last 20 years have seen the emergence of an "audit culture" across society in general (Power 1997) and concerning academia in particular (Strathern 2000a). Underlying this development has been the need and desire for accountability, which is intended to ensure that limited public funds are spent wisely.

Interdisciplinarity is often presented as a response to the increased demand for accountability; it is thought to somehow make knowledge more relevant. One of the most widespread accounts of the recent development is the distinction between Mode-1 and Mode-2 knowledge production. According to *The New Production of Knowledge* (Gibbons et al. 1994), the internally directed mode of knowledge production, "Mode-1 science", has during recent decades been complemented by a demand-driven process, "Mode-2 science". The latter mode attempts to bridge the epistemic gaps that have emerged between disciplines as a result of increasing specialization. The kind of knowledge production associated with Mode-2 is often called "transdisciplinarity" and includes topics that are defined in categories of broader social relevance than found in Mode-1 academia—thus the greater accountability of Mode-2. The suggestion of this view of science policy is that the bulk of resources should be shifted from Mode-1 to Mode-2 knowledge production.

Despite this general consensus, there are divergent views on what constitutes accountable knowledge production, and how exactly interdisciplinarity promotes this goal. In particular, the implications of interdisciplinarity for academic disciplines are seen differently by different parties to the debate. The standard view of interdisciplinarity presumes that the complementarity between disciplinary depth and interdisciplinary breadth is appropriate and to be expected, since interdisciplinary matters are best

addressed by those who can mobilize a range of highly developed expertise (for a critique of this approach, see Fuller 2010). The more radical side of the debate, in contrast, has strongly questioned the legitimacy of academic science and argued for a transition to "post-normal" science (e.g. Funtowicz & Ravetz 1993), where "scientific goals are controlled by political or societal actors", and "scientists' integrity lies not in disinterestedness but in their behavior as stakeholders" (Kunseler 2007, 3-4). In the latter approach, scientific values are displaced by a set of other norms (e.g. Nowotny 2006; Ziman 1994 & 1996 & 2000), and science is understood as a "continuation of politics by other means" (Elzinga 1993).

At either side of the debate, interdisciplinarity is rarely seen as making knowledge more *epistemically accountable*. While ideas of synthetic knowledge, synoptic views, unified science, and other characteristically intellectual values of interdisciplinarity have always been part of the discourse (e.g. Klein 1990 & 1996; Miller 1982), the epistemic benefits of extra- or interdisciplinary verification are not fully addressed in the policy-oriented literature of interdisciplinarity. Many discussions in social epistemology, science policy, and related fields have implied, however, that more extra- or interdisciplinary communication would make science run better (e.g. Campbell 1969; Frodeman & Mitcham 2004; Fuller 1993 & 2000; Gulbenkian Commission 1996; Guston 2000; Weinberg 1962).

According to Fuller (1988), for example, scientific knowledge production has many of the same characteristics as other organized social activities, and from this viewpoint, the autonomy—to the point of isolation—of the disciplines is unwarranted and actually counter-productive. He illustrates this by a simple analogy from business management (Fuller 2002, 33), which suggests that what is good for a discipline is often bad for the entire "business" of producing knowledge. He also cites evidence from the philosophy and history of science to show that the paths taken by disciplinary science may easily become irreversible because of investments made in certain trajectories of intellectual production. He thus defines "reversibility" as the basic indicator of advancement in science, because it is crucial for collective learning and for responding to the changing needs of society (Fuller 2000 & 2002). Similarly, comparative studies in economic sociology (Stark 2009; Whitley 2007), knowledge management (Carlile 2004; Cohen & Levinthan 1990), and even evolutionary biology (Levins 1968) have emphasized the ecological advantages of heterogeneous organizations and the need to understand and manage this heterogeneity (see also Morin 2008).

In the light of these discussions, science might do better if governed more like a single, yet heterarchical, organization. Disciplines, sub-disciplines, etc. can be understood as mutually dependent units of intellectual coordination and control that set the normative and cognitive rules that govern knowledge production and evaluation. Interdisciplinarity, in turn, can be portrayed as the key mechanism of coordination and control *between these units*. It operates as a counterforce to both disciplinary autonomy and the "tyranny" of political or external forces over epistemic values, which have undesired consequences for science: One the one hand, the autonomy of disciplines means subsuming the governance of scientific knowledge production under disciplinary quality control, and science loses sight of overall goals. On the other hand, displacing the power of academic institutions by

the power of markets or political forces, auditing technologies, or other changing fashions runs the risk of destroying the scientific enterprise in the long run (see Ziman 1996). How can interdisciplinarity help science cope with the challenges facing it today? I will next elaborate this question from an epistemic perspective.

2.3 Interdisciplinary accountability as a socio-epistemic mechanism

Reliability is a major epistemic value in science. The search for reliable knowledge is firmly institutionalized in scientific practice; a good example is the pervasive peer review system. As Nowotny and colleagues have argued in their *Re-Thinking Science* (2002; see also Nowotny 2003), however, reliable knowledge is reliable within bounds. In its conventional and limited sense (i.e. Mode-1 science), these bounds embrace a relatively small number of peers; but now, when science has entered the "agora", the boundaries that contain reliable knowledge have been dramatically extended, even abolished. As a result, the constitution of what is considered reliable knowledge in Mode-2 science is more heterogeneous and more ambiguous than ever before.

It is assumed that a shift to Mode-2 knowledge production will improve the reliability of knowledge in contemporary societies. The authors explain this by the logic that the more highly contextualized and infiltrated into social spaces the knowledge, the more reliable it is likely to be, because it remains valid outside the "sterile spaces" created by experimental and theoretical science. They argue that some of the socio-epistemic mechanisms on which scientific reliability is thought ultimately to depend, such as "consensuality" and "consensibility" (see Ziman 1991), can work better if practiced at the level of a wider network of collaborators than within a disciplinary context (Nowotny et al. 2002, 166-178). Thus, knowledge produced under Mode-2 conditions is reliable in terms of relevance for the context in which it arose, and which continues to influence it.

The epistemological core of the Mode-2 argument is based on the idea of social epistemology. Whereas classical epistemology is concerned with the pursuit of truth, and set questions about how justified, true belief can be attained, social epistemology is concerned with the social dimensions of epistemic justification. According to many social epistemologists, the social does not contaminate the normative, or justificatory, dimension of science (Goldman 2010). To the contrary, justificatory reasoning is part of a social practice of challenge and response. Helen Longino, for example, has argued in *The Fate of Knowledge* (2002) that social factors can be incorporated into our definition of knowledge without relinquishing its cognitive rationality. Moreover, social epistemology, such as classical epistemology, is not confined to the description and explanation of science, but can also be seen as a normative enterprise. Like Fuller (e.g. 1988 & 2000), for example, I view the social dimension of knowledge as playing a crucial role in normative considerations about how science should be organized and run.

From the perspective of social epistemology, the transition from Mode-1 to Mode-2 is but one route to more reliable knowledge. It is also based on a contrived dichotomy between disciplinary and transdisciplinary research: While the former is assumed to be

concerned with maintaining disciplinary rigor, the latter is driven by social forces. What this construction remains silent about is the capacity of interdisciplinarity to produce more reliable and relevant knowledge. The reciprocal challenges and responses between intellectual fields, however, presumably improves the reliability and relevance of knowledge in the same way as increased social contextualization, i.e. by broadening the context in which the socio-epistemic mechanisms of consensuality and consensibility operate. If Mode-2 produces a new kind of reliability, interdisciplinary science produces yet another kind of it. In the latter case, intellectual fields would learn from each other, not only from lay groups and other non-academic stakeholders. The resulting knowledge would be more relevant in the sense that it aspires to respond to the expectations of multiple epistemic stakeholders.

As opposed to the disciplinary notion of reliable knowledge, interdisciplinary reliability—related to such notions as "relevance", "robustness" and "field rigor"—may be seen as involving a delicate balance among several and often competing interests and values within science (Frodeman 2010b). This capacity is indispensable to the governance of science in a knowledge society, a major challenge of which is to deal with the complexity and dissonance of knowledge (e.g. Morin 2008). The epistemic power of scientific knowledge will crumble unless we can reconcile with the competing values and claims that coexist in academia. Interdisciplinarity, I argue, has the promise of locally managing the conflicts and discrepancies between disciplinary knowledges. This promise can be delivered due to the broader constituencies involved in interdisciplinary accountability, analogous to the promise of transdisciplinarity in the context of social accountability.

As a normative concept in political science, accountability posits that all relations of authority, inder to be both legitimate and effective, must rest on principles of accountability: those responsible must be answerable or capable of giving an account of their actions (e.g. Dubnic & Frederickson 2011). Power relationships in higher education and research systems, like in other public sector organizations, have in recent decades been influenced by New Public Management: What was once hierarchically governed by state power is now more horizontally governed by multiple stakeholders (e.g. Whitley 2011). The growth of transdisciplinary (Mode-2) and interdisciplinary science can be construed as developments similar to New Public Management, but occurring in the realm of epistemic authority. In both forms of knowledge production, scientific expertise can be seen as a form of delegated authority, distributed according to lines of lateral or horizontal accountability. While transdisciplinarity means that researches act on behalf of democratic publics (cf. Jasanoff 2003), interdisciplinarity means that researchers act on behalf of the scientific community as a whole: they should not be guided by disciplinary interests only, but simultaneously justify their actions to multiple disciplinary constituencies.

Discussions about accountability typically emanate from three fundamental questions: accountability *for what*; accountability *to whom*; and accountability *through which mechanisms* (e.g. Kearns 2011, 199). Disciplines have an in-built accountability of a kind—namely, one that is self-monitoring and epistemological, i.e. having their own standards and theories for how knowledge is made and where it comes from (Strathern 2004, 68). Interdisciplinary research, by definition, does not fall within the remit of any

one discipline, but many. This does not mean that demands for accountability do not apply, but rather that the demands are contingent on more than one discipline. However, what is or is not deemed accountable in each case is a local affair, as the specific meaning of accountability is construed and modified by the intervention of other disciplines. This is where the major epistemic promise of interdisciplinarity lies—in its capacity to intervene in disciplines and change what they do (see Fuller 1993). According to Robert Frodeman (2011, 108), who draws on Wolfgang Krohn (2010), "interdisciplinarity prospers by staying close to cases, expanding a repertoire of skills for dealing with disparate groups in different situations, while resisting the urge for law-like generalizations". This is also why interdisciplinarity appears resistant to epistemological definition and evaluation: it keeps challenging the prevailing epistemological structures.

While interdisciplinary accountability promises improved governance of knowledge, it can also be conceived as a virtue, an end in itself. As a normative condition, "being accountable" means being transparent, taking responsibility for one's actions, and subjecting oneself to scrutiny, control, and guidance (Dubnic & Frederickson 2011). Just as the public accountability of science is a moral stance derived from democratic values, interdisciplinary accountability can be viewed as a norm pertaining to the *ethics of science*: It has been argued that epistemic responsiveness is ethically good in itself (Doucet & Mauthner 2002), and the virtues of interdisciplinarity can be attributed to ethics in one way or another (as in Balsamo & Mitcham 2010). In this sense, interdisciplinary accountability is essential element to the ethical conduct of science; while it is not unlike Mertonian norms (Merton [1942] 1973), it also provides remedies for the fragmentation, narrow-mindedness, complacency, and other "vices" of contemporary academic science. Consequently, interdisciplinary accountability may be virtuous also when it does not actually lead to better research policy, improved governance, or even more reliable knowledge. Emphasizing the latter view does not mean that epistemic values are considered sacrosanct as "scientism" would have it, but that they have a distinctive and valuable role in human culture (Collins 2009 & 2012).

Thus, I suggest that interdisciplinarity has become, for reasons that are not widely addressed in the literature, a pervasive and important element of scientific knowledge production. In particular, while the virtues and benefits of interdisciplinarity have aroused wide interest, they are not usually articulated by referring to epistemic accountability. In the next section, I set my notion of interdisciplinary accountability in relation to the current approaches to the evaluation of interdisciplinary research, and illustrate how it may remedy some problems inherent in those approaches.

2.4 From evaluating interdisciplinarity to interdisciplining evaluation

As discussed in the introduction and in *Article I*, there is little knowledge or consensus on how to evaluate interdisciplinary research, which does not seem to fit in well with the current system for producing scientific knowledge. As a response to this problem, a specific discourse devoted to the evaluation and criteria of interdisciplinary research has

emerged (e.g. *Research Evaluation* 2006). In this, competing positions on interdisciplinarity have led to competing assumptions about quality and how it is best determined (Klein 1996, 211). In *Article I*, I distinguished between three evaluative approaches, which I called "mastering multiple disciplines", "emphasizing integration and synergy", and "critiquing disciplinarity". Each approach defines, implicitly or explicitly, a set of standards against which interdisciplinary efforts are evaluated, and presupposes a context in which their worth is considered.

The evaluative perspectives articulated in *Article I* differ in terms of the extent to which they challenge the disciplinary structure of evaluating knowledge. "Critiquing disciplinarity" is the only one that questions the disciplinary model of intellectual practice—the notion that disciplines (including interdisciplines, as hybrid yet esoteric domains of expertise) have a legitimate authority to define their own goals and standards. Thus, it is a position in line with the idea of interdisciplinary accountability, and is therefore adopted as the overall position of this dissertation. I do not deny the contributions of the other two approaches, or take a radical departure from those discourses, but seek to *shift the focus*: Instead of conforming to the current concept of research quality, interdisciplinarity offers an alternative perspective on how to evaluate it. In doing so, it points out several shortcomings in the disciplinary model of evaluating research.

These shortcomings, and the ways in which the discourse on interdisciplinarity has sought to fix them, can be illustrated with the help of the perspective articulated by Egon Guba and Yvonna Lincoln in *Fourth Generation Evaluation* (1989). In their critical analysis, the authors identify three paradigmatic problems of evaluation as a professional practice: a susceptibility to managerial ideology; a failure to accommodate to value-pluralism (the presumption of a value-consensus); and a commitment to realist ontology. The very same problems, I argue, seem to characterize the disciplinary model of research evaluation, and any variant of this model, including the "mastering multiple disciplines" and "emphasizing integration and synergy" approaches of *Article I*, is insufficient or misleading inasmuch as it fails to resolve these problems. The practical implications of interdisciplinary accountability—and especially a lack thereof—will be clarified in the following pages by applying the critical analysis of Guba and Lincoln to academic research evaluation.

The first paradigmatic problem of evaluation is a *tendency to managerialism*. Following the concept of Guba and Lincoln, this means that evaluations are conducted by the rules set by a closed group of people whose needs the evaluation is supposed to serve. In disciplinary evaluations, this group contains one's peers within the same intellectual tradition. Evaluations are thus closed to inputs from other stakeholder groups, who may have other questions to be answered, other ways of answering, and other interpretations to make. Problems of this tendency are now widely acknowledged in research evaluation, and various ways to open up the peer review process have been debated lively (Frederiksen et al. 2003; Holbrook 2010; Luukkonen 2002). Defining interdisciplinarity as "mastering multiple disciplines" does not question this tendency, but only recasts the people who are deemed eligible to make a judgment; the eligibility is still defined on the basis of a technical mastery of a particular kind of research. This approach tries to ensure

that an appropriate spread of experts is represented in interdisciplinary evaluations, and thereby bring about parity of evaluation outcomes between disciplinary and interdisciplinary research (e.g. National Academy of Sciences 2005). Contributions that emphasize "integration and synergy" as the litmus test of interdisciplinarity, for their part, suggest a more interactive "coaching model", in which evaluation rules are set collaboratively between researchers and reviewers (e.g. Spaapen et al. 2007). Such practices, although empowering particular researchers, may strengthen the tendency towards managerialism by encouraging favoritism between those involved while stifling critical voices from outside (see also Janis 1972). Interdisciplinarity thus becomes a self-justifying practice, such as occurs in disciplinary research.

Another shortcoming in evaluations, closely related to the managerial tendency, is a *failure to accommodate value-pluralism*. Scientific evaluations are dominated by the values and interests particular to the discipline, though the evaluation always affects the values and interests of other disciplines as well. In the current practice of peer review, the concerns of other disciplines are systematically excluded, which is particularly detrimental to interdisciplinary accountability, the robustness of knowledge, and value-pluralism. The "mastering multiple disciplines" approach does acknowledge the pluralism of epistemic cultures, and would incorporate a more diverse set of epistemic norms in the evaluation of interdisciplinary research (e.g. Grigg 1999, 48). However, it takes disciplinary norms as given and immutable, instead of opening them to negotiation and mutual testing. The "emphasizing integration and synergy" approach, in turn, creates a new set of criteria for interdisciplinary research. A number of scholars (Bergmann et al. 2005; Stokols et al. 2003 & 2008 & 2010; Klein 2006 & 2008b; Spaapen et al. 2007) have recently offered concepts and tools for assessing the performance of interdisciplinary efforts with respect to several integrative goals. However, claiming interdisciplinarity as a new genre of expertise in its own right, risks repeating the same problems that have plagued disciplinary knowledge production: insularity, overproduction, and lack of relevance and timeliness (Frodeman 2011; Fuller 1993). Neither approach to interdisciplinary evaluations, therefore, solves the problem of how value differences might be negotiated.

The third, and most profound, paradigmatic problem of evaluation, Guba and Lincoln say, is the *commitment to realist ontology*. Evaluations are typically understood as measurements, descriptions, or judgments concerning the merit of the subject matter at hand, although they are, Guba and Lincoln argue, negotiations about meanings and values. The standard view of academic evaluation puts a premium on meritocratic criteria, against which strengths and weaknesses are evaluated (e.g. Thorngate et al. 2009; Marsh et al. 2008), and even if experienced scholars are sometimes unable to explicitly articulate the criteria, they claim to "know" good research when they see it (see Collins & Evans 2007; Dreyfus & Dreyfus 2005). Therefore, a major concern of the "mastering multiple disciplines" approach is to broaden the evidence base of the evaluation to cover more than one discipline. The "emphasizing integration and synergy" approach, in turn, works towards a better understanding of integrative activities in their own terms. This approch highlights the functions of evaluation for organizational learning, interdisciplinary research performance, and credibility, and its major concern is to develop and disseminate principles of good interdisciplinary practice (e.g. Pohl & Hirsch Hadorn 2007). Absent in

all these views, however, is a critical questioning of the idea of meritocratic criteria in itself, and of the pursuit of incremental improvement of the status quo.

As these flaws seem inherent to the disciplinary model of knowledge production and evaluation, remedies must be sought from an alternative view. While not explicitly discussed in *Article I*, the "critiquing disciplinarity" approach offers an alternative perspective on the very idea of research evaluation. Similarly, some critical evaluation theorists (e.g. Guba & Lincoln 1989; Schwandt 2002), though separate from proponents of the latter approach, imply that research evaluations should not, in the first place, be concerned with determinining the worth of research; instead, evaluators should first be asking what exactly it is they are to determine. The question of *whose* interpretations and values are to be taken into account, and *how* different epistemological positions might be accommodated, becomes paramount. It is in the context of such questions that the attempt to "interdiscipline" academic evaluations can be realized in practice, and the aspects of interdisciplinary accountability be defined. Interdisciplining evaluation means giving voice to representatives from other disciplines, and it therefore releases scientific knowledge production from the constraints of professional criteria. This view is in line with several contributions to interdisciplinary evaluation (e.g. Fuller 2000a & 2002; Laudel 2006; Sarewitz 2000; Weinberg 1962), only few of which, however, have made empirically grounded suggestions for interdisciplining research evaluations in practice.

There are, however, strong pressures driving the institution of peer review toward inter- and transdisciplinarity (see Holbrook 2010). Science agencies have designed their review processes in order to balance the competing values of autonomy and accountability (Holbrook & Frodeman 2011). A consensus prevails that the interdisciplinary aspects of research are best evaluated by a panel of experts from different fields, rather than by any single expert. The rationale for using evaluation panels is twofold: first, to broaden the expertise available for making a judgment, and second, to enable face-to-face deliberation between experts with varying views (e.g. Boix Mansilla et al. 2006; Lamont 2009; Langfeldt 2004). Unlike most other methods of research evaluation, consensus-seeking panel deliberations hold the promise of remedying the paradigmatic flaws in current evaluation practices. However, little is known about the negotiation routines actually used in various evaluation panels or the implications of the routines for interdisciplinary accountability.

3 Methodology

3.1 Research strategy

The empirical part of this dissertation focused on a particular instrument for the governance of science: the evaluation of research proposals. The evaluation of research proposals operates through an accountability mechanism that uses incentives, i.e. research funding, based on account giving: Researchers are expected to be epistemically accountable for the research they propose, and best proposals are rewarded with research grants. The justifying accounts given by researchers for their proposed research are verified by a group of reviewers, who are asked to use their expertise to evaluate the proposals according to a set of meritocratic criteria, and then give their own accounts of their evaluations. I analyzed this double-layered accountability mechanism from an interdisciplinary point of view. In particular, I analyzed how account-giving relationships across disciplines play out in terms of the epistemic content of research proposals as well as the deliberation process on their merits.

The rationale for studying epistemic accountability at this level is that informal routines and conventions—rather than formal rules, norms, or obligations—give accountability its meaning (see Considine 2002). For example, despite the vast array of criteria used to evaluate research quality, studies of researchers' conceptions of quality suggest that good research is something that they "feel" or "know" by virtue of experience, rather than by explicitly following certain norms (Gulbrandsen 2000; Lamont 2009; Langfeldt 2004; Thorngate et al. 2009). Similarly, it is in the researchers' concrete justificatory practices that various features of epistemic accountability are likely to appear.

The concept of "boundary work" (Gieryn 1999) is often used in science studies to illustrate the composite set of claims, activities, and institutional structures that define and protect knowledge practices. In their scholarly work, individuals and groups implicitly create, maintain, and reformulate boundaries between domains (Klein 2008b). In analyzing interdisciplinary activities, I looked at the ways in which epistemic legitimacy is created *despite* disciplines, i.e. by setting aside instead of drawing disciplinary boundaries. As an analytic concept, interdisciplinary accountability is thus understood as the opposite of disciplinary authority.

The empirical setting for the dissertation was a major national research funder, the Academy of Finland (see Table 2). The Academy funds high-quality basic research across all disciplines and fields; it consists of four Research Councils, including Biosciences and Environment, Culture and Society, Natural Sciences and Engineering, and Health. From their several funding mechanisms, I focused on Academy Projects and its evaluation process. A major share of the Academy's research funding—about 30%—is distributed annually through this instrument. As opposed to other forms of funding, this instrument has no specific policy goals, but is targeted to support research that fulfills high academic standards. There are neither particular incentives for interdisciplinary research nor formalized procedures for evaluating interdisciplinary applications (as noted also by Gibbons et al. 2004). The funding decisions for these projects are made primarily on the

basis of peer review statements[5] concerning the research plan and the research environment. Thus, focusing on this particular funding instrument was ideal for analyzing interdisciplinary accountability as a part of the ordinary course of knowledge production and evaluation, that is, in the context of basic research.

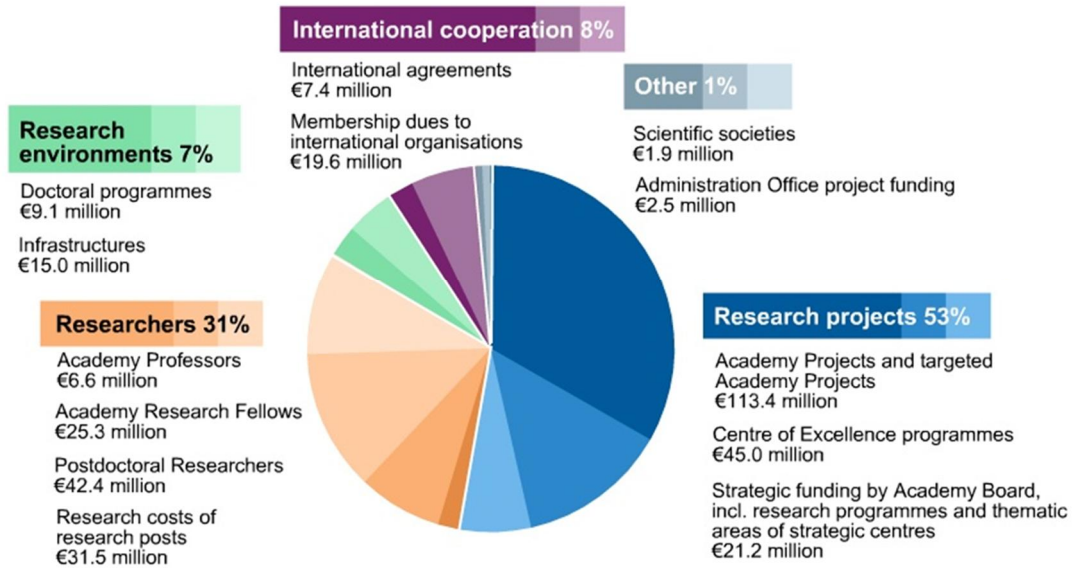
Table 2. *Facts about the Academy of Finland. (Source: Academy of Finland 2012)*

ACADEMY OF FINLAND

Year of foundation: 1948 (the current form of operation founded in 1970)

Annual budget for research funding: 327 euros (year 2012)

Research funding distribution (year 2011):



Total €341 million

The analysis draws from the two "layers" of account giving in the context of the Academy Projects. First, I analyzed the content of successful research proposals, including discussions and assumptions concerning disciplines and interdisciplinarity (*Article II*). Second, I studied peer review deliberations, the process through which expert panels consider the merits of research proposals (*Articles III-IV*). In both studies, the analytical

[5] *The peer review model used by the Academy is similar to that used in most other European countries, and it is described in Articles III-IV.*

focus was on the various ways scholars justify and account for their actions among and across disciplines. However, as research proposals and peer review deliberations are such different objects of investigation, the analyses were conducted quite differently.

3.2 Data and methods

The empirical data were collected during 2004-2007. The data gathering process occurred in two rounds, the first of which was part of a consultancy project commissioned by the Academy of Finland (*The Academy of Finland and the Promotion of Interdisciplinary Research*—AFIR, see Bruun, Hukkinen et al. 2005), and the second aimed at exploring one of the undescribed areas that emerged during the first round. The data consists of a diverse set of (1) research proposals, (2) evaluation reports, (3) interviews with funding officers, and (4) interviews with peer reviewers. Each data category is addressed below, and all data are summarized in Table 3.

1. Research proposals. Two sets of research proposals were included in the study, but the intensity of their analysis varied notably: while some were read several times, others were only glanced over for topic and abstract. As part of the AFIR project, I analyzed a sample of research proposals that had received funding in 1997 and 2000 through Academy Projects (n=266; 59% of the funded proposals for those years). This sample included research proposals across all research fields. Of that set, approximately one third (n=106) was categorized as interdisciplinary to some extent, and those proposals were analyzed more carefully. The other round of data gathering focused on another set of proposals, submitted in 2007. This set consisted of 109 proposals from environmental and social sciences. Included were all proposals evaluated by the four peer review panels selected for this study (see *Articles III-IV*). The latter proposals were read mainly in order to understand the evaluation reports and to prepare for the interviews with the peer reviewers. The most careful reading was given to those proposals that received conflicting evaluations from the peer reviewers. In addition, proposals that had interdisciplinary characteristics were read more closely than others. The research proposals consisted of several documents, but the analysis focused on the contents of the research plans only, since that is where the project was described (see Appendix 1).

2. Evaluation reports. Evaluation reports on the set of 109 research proposals (see above) were analyzed in the second round of data gathering. This material consisted of two kinds of reports: the preliminary reviews and scores given to the proposals by individual panelists prior to the meeting, and the final reviews and scores as defined collectively by each panel as a group. For each proposal, there were usually three different evaluation reports: two informal reports from the preliminary reviewers, and one final report. In addition, two chairperson's reports on panel meetings were included in the study. As was the case with the research proposals, the evaluation reports were analyzed in order to prepare for the interviews with the evaluators themselves. Here, again, the most careful analysis focused on the preliminary reviews that indicated different views on a proposal, as well as on the respective consensus reviews. A number of evaluation reports were used as secondary material in the first round of data gathering, too, but their role was

to help categorize the proposals, and only a fraction of them was actually available. The evaluation reports were in the form described in Appendix 2.

3. Interviews with funding officers. A total of ten semi-structured interviews were conducted with the funding officers. In the first round, I interviewed eight funding officers in order to find out the formal procedure of evaluation and to ask for the officers' views and experiences of the evaluation of interdisciplinary proposals as compared to disciplinary proposals (see Appendix 3). I selected two experienced officers from each of the four research councils for an interview. In the second round, I interviewed two funding officers who convened and attended the four panel deliberations selected for this study (see Appendix 4). The interviews lasted approximately one hour, and they were recorded and transcribed in part.

4. Interviews with peer reviewers. Data collection in the second round involved phone interviews with eighteen (out of twenty-seven) panel members who served on the four peer review panels. The selection of interviewees included the majority of those panelists who were willing to be interviewed within a few weeks after the panel meeting. While the schedule of the interviews was the same for all (see Appendix 5), the substantive content of the interviews was tailored to each panelist. During the interviews, the panelists were asked: to profile their personal expertise as well as the collective expertise of their panel; to explain what happened in selected controversial cases; to describe the arguments they themselves made about a range of proposals; to contrast their own arguments with those of other panelists; and to consider the meaning of the deliberation process in general. The interviews lasted approximately one hour, and they were recorded and transcribed fully. The transcripts were analyzed with the help of the qualitative analysis software Atlas.ti.

Table 3. *Research data used in this dissertation.*

	Research proposals	**Evaluation reports**	**Interviews with funding officers**	**Interviews with peer reviewers**
Round I (2004–5)	n = 266 Primary material for categorizing interdisciplinarity	n = Indefinite Secondary material for interpreting proposals	n = 8 Secondary material for learning about evaluations in the Academy	--
Round II (2007)	n = 109 Secondary material for interpreting deliberations	n = approx. 329 Primary material for identifying disagreements	n = 2 Secondary material for interpreting deliberations	n = 18 Primary material for analyzing deliberations

Data analysis in the first round of the study was an exercise to categorize research proposals according to their interdisciplinary properties. The empirical unit of analysis was a research proposal that had received funding from the Academy Projects fund. Content analysis of the proposals was conducted to investigate where the research stood in the system of disciplines, in what way the disciplines were intended to be brought together, and what the goals of the research were. The research proposals were analyzed with a classification scheme developed on the basis of a conceptual and practical discussion on interdisciplinary research (for a description of this framework and how it was used, see *Article II* and Appendix 6). The analysis thus followed mainly a top-down approach, where the typology of interdisciplinary research was used as a theoretical apparatus to analyze the content of research proposals. From the standpoint of the dissertation as a whole, however, the analysis can be characterized as a bottom-up attempt to investigate what constitutes interdisciplinary account giving, and how it can be operationalized. In relation to the latter goal, the analysis of the interdisciplinary aspects of research proposals provides empirical evidence.

Data analysis in the second round of the study, in turn, focused on interdisciplinary accountability in the making of evaluations, i.e. as it unfolds in the deliberation among peer reviewers. Here the unit of analysis was a deliberation process by a peer review panel evaluating a sub-set of the Academy Project proposals. The aim was to explore the ways in which reviewers deliberating in panels of varying disciplinary compositions bridge their epistemological differences and negotiate criteria of what constitutes quality (for the specific design of that round of the study, see *Articles III-IV*). The intersubjective perspective adopted here (see e.g. Garfinkel 1967) illuminates how reviewers make sense of their judgments in the presence of several and sometimes competing disciplinary regimes. I analyzed the informal rules of deliberation and focused on the emerging accountability relationships between the reviewers. The idea of the analysis was to study peer deliberations as indeterminate situations, where the traditions, conventions and institutionalized customs are not left unexamined, but instead opened up for reflection by the actors themselves (see Stark 2009, 32).

4 Research findings

In this section, I discuss my key findings concerning interdisciplinary accountability in the evaluation of research proposals. In particular, I discuss options to demonstrate, validate, and strengthen interdisciplinary accountability within the chosen funding instrument by reconsidering the concepts and practices on which it is based. The idea is to find ways to encourage: (1) researchers to write their proposals with an eye toward convincing not only experts in their own field, but also other researchers for whom the given research may count in one way or another; and (2) reviewers to consider the quality of research proposals beyond any particular field of expertise. Thus, the findings can help in designing the proposal evaluation process so that the accountability relationships across disciplines are routinely considered by both researchers and reviewers, instead of letting disciplinary norms tacitly govern the process.

As mentioned earlier, I drew on two empirical sources of epistemic accountability: (1) the content of research proposals; and (2) the deliberation procedure through which peer review panelists evaluated the proposals. The findings are reported in two parts. Drawing mainly on *Article II*, Section 4.1 responds to the first research question: How is interdisciplinary accountability to be conceptualized in research proposals? Here, I delineate a framework for analyzing the key aspects of interdisciplinary accountability in research proposals. Drawing on *Articles III-IV,* Sections 4.2 and 4.3, in turn, respond to the second research question: How can peer review facilitate interdisciplinary accountability? In these sections, I analyze the practices and design of peer review deliberation, which is the major mechanism through which epistemic accountability of research has traditionally been maintained and fostered.

4.1 Interdisciplinary accountability in research proposals

The principles of interdisciplinary accountability are necessarily complex, as interdisciplinary knowledge is based on lateral patterns of epistemic authority. While the standard model of scientific evaluation relies on the inbuilt accountability of disciplines, interdisciplinary evaluation would consider the constituents of accountability as a more open-ended question. According to the concepts of evaluation theory and practice, the stakeholders or *beneficiaries* of accountability are those whom the evaluation is supposed to serve. The contents of accountability, in turn, usually refer to *goal* accountability, *process* accountability, and *outcome* accountability[6] (Alkin 1972 & 2004). For an interdisciplinary interpretation of these accountabilities in the evaluation of research proposals, I suggest using the typology and indicators of interdisciplinary research articulated in *Article II*.

[6] *As this research focuses on the evaluation of research proposals, i.e., the ex ante evaluation of research, outcome accountability will not be discussed here.*

Classification schemes such as the one presented enable us to think about, categorize, and evaluate research in a way that is an alternative to the system of disciplines. As opposed to the disciplinary classification of knowledge that demarcates domains of specialized inquiry, interdisciplinary classification schemes differentiate between "forms of disciplinary interaction, motivations for teaching and research, degrees of integration and scope, modes of interaction, and organizational structures" (Klein 2010, 15). They direct attention to interdisciplinary aspects of research and away from the disciplinary aspects (Klein 2008b; see also Bowker & Star 1999). What is novel in the suggested typology in relation to earlier classification schemes is that it effectively reveals the accountability structure of interdisciplinary research—that is, to whom and for what interdisciplinary research is epistemically accountable.

Drawing on an extensive review of the literature on the epistemic characteristics of interdisciplinarity, as well as on a content analysis of 266 research proposals, the typology illustrates how to conceptualize interdisciplinary accountability in research proposals in terms of the following aspects: (1) the *beneficiaries* of accountability, who are defined by the "scope of interdisciplinarity", i.e. where the proposal is located in the system of disciplines; (2) *goal* accountability, defined by the "type of goals", i.e. what functions the proposed research is intended serve; and (3) *process* accountability, defined by the "type of interdisciplinarity", i.e. how it is conducted. The constituent elements of interdisciplinary accountability in each case, I suggest, can be specified by examining where the proposal stands in terms of these three dimensions. Further, I claim, these considerations should be a *prerequisite for every evaluative act.*

An operational definition of each aspect of accountability in the evaluation of research proposals, as well as an interdisciplinary interpretation of these aspects, is presented in Table 4. However, as the classification scheme in *Article II* was originally designed for the analysis of *interdisciplinarity* in research proposals, not for the analysis of *interdisciplinary accountability*, some conceptual remarks are in order here before going into the definitions. In particular, some of the original definitions may appear too static, even ill-fitted, for the perspective of interdisciplinary accountability construed in this dissertation summary. In *Article II* I and my co-authors distinguished between "discipline" and "field" (or specialty), the former being an institutional concept, the latter an intellectual concept, and defined interdisciplinarity in relation to the latter. Intellectual fields, however, also have institutional functions. They organize and govern intellectual production just like disciplines, though in a more informal and fluid manner—they can be construed as *cognitive institutions* (see e.g. Scott 1995). And it is exactly these institutional functions that interdisciplinarity challenges.

Thus, the distinction between "discipline" and "field" indicates a difference of *scale* rather than a difference of underlying mechanisms. Here, the notion of a "disciplinary form" (Cambrosio & Keating 1983) is helpful in analyzing cognitive institutions at different scales. It implies, for example, that if biology can function as a discipline (that is, to take a disciplinary form and hold a disciplinary stake) with respect to the scientific enterprise as a whole, then biology may itself constitute an overarching scientific enterprise with respect to which a smaller category, e.g. chronobiology (in Cambrosio and Keating's example), can function as a discipline. Thus, there is no need to respond to

questions intended to determine, once and for all, whether a given scientific practice does indeed constitute a discipline. Following this multi-scalar or relative notion of discipline (see also Abbott 2001), "interdisciplinarity" can also be given a relative meaning that is not static or exhaustively descriptive: what it designates is not a particular form of knowledge production (as in e.g. Repko 2008), but a deliberate yet open-ended pursuit of intellectual change.

That said, the categories provided in *Article II* do remain valid and useful operational measures of interdisciplinary accountability in research proposals. Below, I illustrate how the concepts can help in the designing and evaluating of interdisciplinary research proposals.

Table 4. *A horizontal or interdisciplinary interpretation of accountability in the evaluation of research proposals (MD=multidisciplinarity, ID=interdisciplinarity).*

	Beneficiaries of accountability	**Goal accountability**	**Process & outcome accountability**
Operational definition	The target audiences of research	The functions and benefits of research	Procedures for accomplishing the goals
Respective dimension of interdisciplinarity	"Where"/"who": The scope of interdisciplinarity	"Why": The type of goals	"How": The type of interdisciplinarity
Suggested subcategories	Disciplinary scope Narrow interdisciplinarity Broad interdisciplinarity	Disciplinary goals Epistemic goals Instrumental goals Mixed goals	Disciplinary procedures Encyclopedic MD Contextualizing MD Composite MD Empirical ID Methodological ID Theoretical ID

4.1.1 Beneficiaries of accountability

As indicated in the introduction, a central issue in evaluating interdisciplinary research concerns the intellectual context from which to draw the standards for assessing quality. Obviously, since there are no universal criteria for quality, each account may seem

justified and make sense to particular audiences. In terms of accountability, the issue concerns the relevant others or beneficiaries to whom the research is epistemically accountable. Therefore, in accounting for interdisciplinary interactions, the first question is "Where is the given interaction located in the system of disciplines?" i.e. what are the disciplinary forms between which the activity occurs. Using the terminology suggested in *Article II*, this property of research is the "scope of interdisciplinarity". The discourse of interdisciplinarity has theorized about the implications, variations, and analytical distinctions of the scope (e.g. Kelly 1996; Klein 2005; Porter and Rossini 1984), but I argue that scope also defines the appropriate context for the evaluation, i.e. the normative, interpretative, and authoritative grounds on which a given research project operates.

However, as discussed in *Article II*, specifying the scope of interdisciplinarity is, by nature, quite fraught with ambiguity and may ultimately be simply a matter of preference: Interdisciplinarity can occur at different "levels" of the nested system of disciplines (Kastenhofer et al. 2002; Cambrosio & Keating 1983; Ziman 1997). On this basis, *Article II* made a simple distinction between narrow and broad interdisciplinarity. The target audience of research is different depending on what institutional boundaries are crossed. The broader the potential audience, I suggest, the broader is the potential significance and relevance of the research, but also the variability between stakeholders, whose epistemic expectations are to be responded to. Whether sympathetic or critical in tone, incremental or revolutionary in effect, interdisciplinary research is assumed to make sense to those regarding whom it operates, criticizes, comments, or synthesizes. Disciplinary research, in contrast, is assumed to be accountable to disciplinary stakeholders only.

As an example, one Academy Projects proposal concerned the historical changes of community life in Southwest Finland. The project proposed to analyze archaeological material as well as natural scientific evidence (e.g. DNA structures and heavy metal concentrations) along with materials from traditional historical archives. The relevant stakeholders of epistemic accountability in this case included anthropologists, archaeologists, natural scientists, and historians, all of whom would need to negotiate together for an appropriate evaluation of this proposal.

4.1.2 Goal accountability

Another fundamental aspect in evaluating interdisciplinary research is the rationale behind it, i.e. why an interdisciplinary research effort is performed in the first place. This aspect indicates the ultimate goal for which a research effort is to be held accountable. Obviously, an interdisciplinary approach is a means to contribute to some desired outcome, not an end in itself. Therefore, goal accountability can be defined by analyzing what functions a given research serves and who will benefit from it (see Burawoy 2005). Using the terminology suggested in *Article II*, this property of interdisciplinarity is called the "type of goals". In the literature on the subject, the variability of interdisciplinary goals is highlighted mainly because the stated goals feature prominently in how the performance is evaluated (e.g. Klein 2008a; Langfeldt 2006; OECD 1982; Salter & Hearn 1996). For example, researchers engaged in pragmatic problem solving or product development

reasonably place a high premium on viability, workability, and impact, while researchers investigating ways to understand multidimensional phenomena typically strive for new levels of comprehensiveness, careful description, and empirical grounding (Klein 2008a). I maintain, however, that the types of goals need to be scrutinized not only in terms of whether and how they can be achieved ("outcome accountability" in Alkin 1972), but in terms of subjecting the goals themselves to evaluation ("goal accountability" in Alkin 1972).

Following the discourse on interdisciplinary research, I distinguish between epistemic, instrumental, and mixed goals. While this classification is very coarse, and more sophisticated typologies do exist (e.g. Burawoy 2005, applied by Mollinga 2008, 9), it does illustrate that, beyond a researcher's specific goals, there are also broader functions and benefits of research to be scrutinized. In particular, such scrutiny is ultimately relational: Given the increasing scarcity of research funding, the urgency of certain problems over others, and the uneven distribution of benefits (and potential harms) of a research project, all goals are not equal. The demands for disciplinary knowledge are not without limit, as all goals cannot be pursued (Frodeman 2011; Fuller 1993). Interdisciplinary accountability for research goals can be differentiated from disciplinary accountability on the basis of the accounts given as to why a particular problem, question, or topic is worth investigating and does not fall within the remit of any single discipline. In terms of goals, then, interdisciplinary research is usually held accountable in a way disciplinary research seldom is—the latter being self-justifying in the sense that responsiveness to the problems within the discipline becomes an end in itself. Interdisciplinary research typically prospers by pursuing very relevant, timely, and reasonable goals, which often diverge from the existing lines of research.

In the example mentioned above, the proposed research was oriented towards broad epistemic goals: its purpose was to develop theories and methods of archaeological research on communities beyond a single disciplinary canon. Goal accountability in that case means that advancing these particular goals was regarded as the most reasonable and justified way to employ the given researchers and to invest the given amount of research funding. The beneficiaries of accountability would be advised to consider, among other things, priorities between investments in solving the conceptual issues above, weighing the proposed funding against investing in other, competing, conceptual goals or perhaps more pragmatic research endeavors.

4.1.3 Process accountability

The third fundamental aspect in evaluating interdisciplinary research concerns the process by which interdisciplinary interaction is accomplished. Just as there is no universal scientific method, there will never be a universal interdisciplinary method (e.g. Frodeman 2010a). From the perspective of process accountability, it is thus necessary to specify and justify the way in which interdisciplinary research is done. Using the terminology of *Article II*, this is the "type of interdisciplinarity". The concept is concerned with specifying the procedures through which intellectual resources across fields are employed

to accomplish the given goals (as in, e.g., Boix Mansilla 2010). As such, it can occur in various ways, tends to be asymmetrical in terms of cognitive transfer, and only rarely leads to integration (e.g. Boden 1999; Boix Mansilla 2006; Bruun & Toppinen 2004; Jeffrey 2003; Klein 1990 & 2010; Lattuca 2001; Mollinga 2008; Rafols 2007). Discussion about this aspect has indeed had the most profound effect on how we understand the phenomenon of interdisciplinarity. For example, the widely used distinction between "multidisciplinarity", "interdisciplinarity", and "transdisciplinarity" is based on an assumption of different degrees or types of knowledge integration. However, the distinction that I suggest, i.e. disciplinary vs. various types of interdisciplinary research, is particularly useful for evaluating interdisciplinarity as a methodological process.

Scholarly research is, by definition, epistemologically accountable for its procedures of acquiring knowledge. My analysis of the various types of interdisciplinarity illustrates how the procedures used may be accounted for in more than one way. Knowledge resources—such as facts, concepts, data, methods, and theories—originally produced by one discipline can be mobilized in new intellectual settings and translated into new epistemic entities (e.g. Bruun, Langlais et al. 2005; Nikitina 2005; Palmer 2001). During such operations, procedural standards and expected outcomes are necessarily reconsidered. I distinguish between encyclopedic, contextualizing, and composite procedures, which tend to be multidisciplinary and coordinative, and empirical, methodological, and theoretical procedures, which tend to be interdisciplinary and integrative. The categories are not mutually exclusive, and a research proposal can represent multiple types at the same time.[7] Interdisciplinary process accountability posits that disciplinary procedures, like disciplinary goals, may be subject to interdisciplinary considerations. From this viewpoint, interdisciplinary procedures for acquiring knowledge will deliver more robust results than disciplinary procedures, assuming that the multiple commitments to process accountability are met.

The Academy Projects proposal, mentioned above, represented two types of interdisciplinary process: empirical and methodological. Interdisciplinary process accountability presumes that the particular combination of evidentiary materials proposed—archaeological material, natural scientific analyses, and historical records—and the methods through which the data sets were to be interpreted and interconnected can provide a reasonable and appropriate basis for research towards understanding the community life changes in the given historical context. These aspects of the proposal would thus be critical negotiation points in evaluating its process accountability.

4.2 Accountability through the customary rules of evaluation

As peer review is the major mechanism through which epistemic accountability in knowledge production is demonstrated and fostered, it is essential to consider options to adjust this mechanism to operate in a more interdisciplinary fashion. To this end, this

[7] *For the original analysis, however, specific criteria were developed to place each proposal in a single category; see Appendix 6.*

dissertation has analyzed the internal functioning of the peer review process through which proposals for the Academy Projects are evaluated. The analysis captures how evaluation unfolds, especially how the disciplinary and interdisciplinary lenses of evaluators, as disclosed and discussed in the deliberation, affect the evaluation process.

A central finding of this analysis concerns the customary rules that peer review panelists follow in evaluating the quality of research proposals (*Article III*). These are intersubjective rules that guide panel deliberations without being formally spelled out. Panelists cannot always articulate these rules, as they often take them for granted. But it is by adhering to such rules that evaluators are able to bridge their epistemological differences and perform the task of evaluating, while maintaining their belief that their evaluation is legitimate (see Lamont 2009; Mallard et al. 2009). *Article III* discusses such customary rules as (1) deferring to expertise and respecting disciplinary sovereignty; (2) pragmatic use of alliances and strategic voting; (3) promoting the principles of methodological pluralism and cognitive contextualism; and (4) limiting idiosyncratic tastes and self-reproduction. Our findings on these practices contribute in three ways to the second research question: How can peer review facilitate interdisciplinary accountability?

First, the prevalence and salience of customary rules in itself offers an analytical perspective that regards evaluation as more about negotiating meanings and values than about measurement, description, or judgment (see Section 2.4). Instead of exploring some "fundamental" aspects of defining quality, the analysis in *Article III* describes peer deliberation as an attempt at communicative rationality; participants aim at a kind of Habermasian ideal speech situation by following the customary rules of fairness. The given analytical perspective stands for the indexicality of evaluation, or the situatedness of any appraisal in a particular context of meaning-making. It thus sets the stage for thinking about evaluations as situationally shaped constructions, unfolding through the evaluators' interactions, and linked to the local context within which they are formed and to which they refer (as in Lamont 2009). Through these customary procedures, disciplinary norms are, indeed, subjected to pragmatic considerations involving fairness and appropriateness. No criterion of quality is valid unless it is first deemed appropriate through fair negotiation, i.e. by following the customary rules of fairness.

Second, some of the most central customary rules, originally observed by Michèle Lamont in her ethnographic study of multidisciplinary funding competitions, *How Professors Think* (2009, Ch. 4), are clearly at odds with interdisciplinary accountability. *Articles III-IV* illustrate that "deferring to expertise and respecting disciplinary sovereignty" is precisely the attitude that discourages experts from making evaluative contributions possibly infringing on each other's intellectual turf, as insights into research in other reviewers' territories are deliberately muted. Similarly, the "principle of methodological pluralism and cognitive contextualism" is shown to implicitly prevent reviewers from challenging other methodological or disciplinary traditions, and lead them to abandon critical appraisal of a research proposal only because it represents a different genre. Customary rules such as these, therefore, tend to legitimate disciplinary authority in concrete evaluation situations, even when the clash of cognitive frames is evident and could, in principle, be openly contemplated. The findings thus explain why it is not always enough to include experts from various fields to collectively evaluate proposals; their

informal practices are likely to keep disciplinary norms "sacred". At the same time, some other customary rules are essential for maintaining epistemic accountability across disciplines. Among them is "limiting idiosyncratic tastes and self-reproduction", that is, subordinating one's personal preferences to more neutral criteria of evaluation.

Third, the comparison of panel deliberations shows that customary rules vary to some extent across settings. This is so for at least three reasons. First, there are discipline-specific practices that researchers are socialized into early on. Modes of evaluation in the social sciences and humanities, on the one hand, and in natural sciences, on the other hand, are clearly different. Second, practices emerge from the dynamics and exigencies of particular intersubjective contexts. The consensual practices of disciplinary panels, for example, differ from those of multidisciplinary panels. Third, practices conform to the formal rules and evaluative techniques imposed by the funding agency. Comparative ranking of proposals in relation to each other produces different behavior than the instruction to rate each proposal on a more abstract scale according to its intrinsic strengths and weaknesses. Due to these (and probably many other) variations, disciplinary criteria of evaluation are challenged more often in some settings than in others.

The findings illustrate that "deferring to expertise and respecting disciplinary sovereignty" is a salient rule in *multidisciplinary* competitions, where panels are composed of distinct experts from different fields. In those settings, this deferential attitude is essential to the collective belief that the process is fair, as it is an efficient way to set aside disciplinary prejudices against others' criteria (Lamont 2009, 135). There is clearly less deference shown in *disciplinary* panels where the specialties of panelists more often overlap. In these panels, arguments occur more explicitly between alternative perspectives. There is also less respect for disciplinary sovereignty in *less specialized* panels concerned with topics that are of interest to wider audiences. In such panels, there is often explicit reference to non-expert opinion as well as to the role of intuition and learning in grounding decision-making. The extent to which panelists defer to distinguished disciplinary expertise and respect disciplinary sovereignty rather than engage in deliberative forms of interdisciplinary accountability therefore seems to be contingent on the particular intersubjective context.

Promoting the principles of "methodological pluralism and cognitive contextualism"—evaluating proposals according to the standards of the discipline of the applicant—was found to be more salient in *humanities and social science* panels than in *natural science* panels. In the latter panels, disciplinary identities may be unified around the notion of scientific consensus, including a shared definition of the indicators of quality. Also panels composed of *generalists*, instead of *specialist experts*, are less favorable to pluralism and contextualism, more often relying on a general matrix of comparison to assess seemingly incommensurable proposals. Promoting the principles of methodological pluralism and cognitive contextualism was thus found to be partly an internalized convention of humanities scholars and social scientists, and partly an emergent practice among specialists.

4.3 Designing accountabilities between peer reviewers

The findings above suggest that interdisciplinary accountability can, indeed, be facilitated by peer review, and that a better understanding of the customary rules of evaluation is necessary for this goal. We cannot control the outcomes of peer review (see Hirschauer 2010), but we can control some of the conditions in which reviewers negotiate. What distinguishes the evaluations academics perform on grant panels from those that occur in departmental evaluations, for example, is *context*. As observed by Lamont (2009, 111) panelists do not engage with each other on a sustained basis, nor do they spend their professional lives with the researchers they evaluate. Thus, while academic culture may not encourage interdisciplinary accountability (see Abbott 2001; Whitley 1984), reviewers across all fields are likely to be exposed to the situational variants of evaluation. One such variant concerns the disciplinary composition of reviewers in a panel.

The comparison of panel deliberations between more and less heterogeneous panels in *Article IV* suggests that the disciplinary composition of a panel creates its own particular sphere of socio-epistemic control and reciprocal accountability, which influences the panelists' tendency to build on and criticize each others' judgments. While peer reviewers who conduct evaluations as individuals are practically unaccountable to each other for their judgments, a new sphere of accountability arises between peer reviewers in the panel setting: experts judge each other's standards and behavior as much as they judge the proposals (see Hirschauer 2010; Lamont 2009). This opens up the possibility to strengthen interdisciplinary accountability by considering the design of peer review panels. Thus, besides including "interdisciplinarity" as a formal criterion for evaluation, research funders can encourage interdisciplinary accountability through indirectly influencing the account-giving relationships between reviewers.

As illustrated in *Article IV*, the extent to which panelists hold themselves accountable to each other for their evaluative behavior, as well as the particular way they do so, are likely to vary according to the scope, composition, and distribution of expertise in a panel. I have demonstrated this by analyzing and comparing the discursive practices through which the members of selected panel deliberations related their expertise and criteria to those of the other members, and by identifying precisely the types of evaluative contributions individual panelists made to the panel's collective judgment. The types of *bridging expertise* practices included allocation, pooling, integration, and generalization. The different practices for *agreeing on proposal ratings*, in turn, included equalization, calibration, contextualization, and commensuration. These distinctions are not claimed to encompass all kinds of interdisciplinary alignment. Rather the idea was to illustrate some of the differences that do exist (see *Article IV*).

I argue that such differences are not random, but result from the panelists' attempt to respond to each other's expectations. The requirement to negotiate one's decisions with other panel members necessarily influences how a panelist engages in the evaluation, since, depending on the others' standards and behavior, some aspects of one's own expertise gain strength while others are left aside. Thus, the matter of for whom a reviewer presents her judgment or in front of what audience her arguments are debated is not without significance. What proposals are on the panel's table also matters, since this

creates a local set of comparables, and thus affects not only one's own judgment on a proposal, but also others' reactions to one's appraisal. I suggest that through temporary accountability relationships, panelists both restrict as well as amplify each other's judgments on proposals. Perceived accountabilities between panel members are likely to vary according to the internalized evaluative norms that prevail in the field (e.g. humanities versus natural sciences), but the composition of a panel has a detectable influence as well.

I also suggest that the *ad hoc* responsibilities arising between panel members tend to create an environment of accountability that, though erected on temporary basis, is likely to hold. This is so also in terms of disciplinary and interdisciplinary considerations. Despite the "evolution" that occurs through the deliberation process, a panel that develops good interdisciplinary dynamics does not, at the same time, allow much disciplinary discretion to individual panel members. As explicated by Edward Hackett and Daryl Chubin (2003), peer review does many things and serves many purposes, but it cannot simultaneously deliver on all things equally well. When interdisciplinary considerations are given priority, at least two important factors stand out.

One concerns the *mix of experts* in a panel, in terms of the overlap in the panelists' competencies. There is a continuum between "disciplinary" and "multidisciplinary" panels, with varying degrees of competence overlap. High overlap in disciplinary panels may induce a relatively tight disciplinary control between participants, including checking each others' appraisals in terms of their validity within the discipline as well as their consistency with other panelists' criteria. A multidisciplinary design, in turn, seems to create a shared sense among panelists that they are accountable for their judgments to a number of different disciplinary communities. At the very least, this leads panelists to explain their standards to each other, but not necessarily to use any particular set of standards or to be mutually consistent in their appraisals. Furthermore, multiple accountabilities may encourage panelists to argue their views by drawing also on sources outside their own particular fields of knowledge, and, ideally, to debate the strengths and weaknesses of entire disciplinary approaches from the perspective of a particular problem. Too little overlap between panelists' competencies, however, may cause them to forgo interdisciplinary dialog and instead divide areas of responsibility between the panel members—i.e. claim and respect others' claims to separate areas of intellectual turf.

Another important factor concerns the selection of panelists in terms of their *degree of expertise in a specialty field*. There is a continuum between "specialist" (including specialists of both disciplinary and "hybrid" fields) and "interdisciplinary" (i.e., generalist) panelists. Interdisciplinary panelists are, from the start, capable of understanding, judging, and comparing a wide variety of proposals without having fully developed expertise on the subject matter. The findings suggest that such panelists also effectively use each others' experience and views as source material for making their own judgments. They also usually accept and operate with multiple epistemic regimes. Rather than keeping a watch on one another's appraisals for consistency with disciplinary authority or appropriateness, such panelists explicitly form their judgments on the basis of the dialog itself, which allows for flexibility and for the individual participants to develop a shared sense of what defines the particular merits of each case.

5 Discussion

5.1 The concept of interdisciplinarity revisited

Interdisciplinarity is in some discourses defined essentially as a critique of the current authority and autonomy of disciplines to govern their knowledge production (e.g. Frodeman 2010a; Fuller 1993; see *Article I*). Like those discourses, this dissertation's vision of knowledge production would replace the disciplinary model of expertise with one based on interdisciplinary reflection, communication, and critique. Thus, while interdisciplinary research defies the existing categories and criteria of much research evaluation, interdisciplinarity may itself be used as a criterion to evaluate and classify research. Thus, interdisciplinarity may play a role in restructuring knowledge production (see e.g. Klein 2008b) and in relieving it of burden of discipline-internal standards. These functions of interdisciplinarity, I suggest, are best grasped with the concept of interdisciplinary accountability, which shifts focus from defining interdisciplinary activities to *redefining disciplinary activities*.

Underlying such a reorientation is the idea that the real challenge of producing—and evaluating—relevant and reliable knowledge is to navigate through different and often competing epistemological regimes. Obedience to a single disciplinary regime is but one choice, and as such, subjected to interdisciplinary consideration. This point is an important contribution to the discourse on interdisciplinary research evaluation, because it avoids the conservative implications of some earlier contributions. In particular, highlighting consistency with antecedent disciplines as a central criterion of interdisciplinarity (as in Boix Mansilla 2006) seems to undermine any truly interdisciplinary critique. Highlighting interdisciplinary accountability, in contrast, implies that it is disciplinary practice that needs to conform to interdisciplinary practice, not vice versa. Each interdisciplinary encounter is, at the same time, a local field test of disciplinary criteria.

In light of the theoretical discussions in Section 2, a continuous struggle between evaluative standards is necessary for accountability in the governance of knowledge production. Interdisciplinary research is only one case in point. Due to being situated on the borders between disciplines, interdisciplinary research is *obliged to actively* search for an audience, consider what is worth investigating, and struggle with determining the norms of good conduct. If we acknowledge that the institutional boundaries between disciplines or fields do not only reflect a functional division of cognitive labor, but indicate the tendency of knowledge practices to develop into self-legitimating "disciplinary forms" (Cambrosio & Keating 1983), then disciplinary communities should routinely be held accountable to each other. This is increasingly the case in the current era of the knowledge society, characterized by uncertainty and dissonance of knowledge, on the one hand, and demands for public accountability, on the other.

5.2 Account-giving relationships in quality control reconsidered

My analysis of interdisciplinary accountability illustrates how account-giving relationships are emerging in quality control. While disciplinary quality control operates within a given epistemological regime, where the rules of the production, interpretation, and legitimation of knowledge are set by a single discipline, interdisciplinary quality control operates in a contingent epistemological environment. In such an environment, researchers and reviewers face the challenge of multiple accountabilities, and each act of account giving is an attempt to explain and justify research from the particular web of intellectual relationships (analogically with political accountability, see Considine 2002; Romzek 2011) that interdisciplinary interaction brings about. From the vantage point of interdisciplinary accountability, disciplines offer useful regimes for the acquirement and certification of knowledge, but such regimes are prone to partiality. An interdisciplinary research effort is accountable to multiple constituencies, and therefore it is always justified or not justified on a provisional basis. I argue that this should be considered as a major strength, not a deficiency.

The findings of this study illustrate actual ways of coordinating, integrating, and negotiating disciplinary regimes in the the pursuit of high-quality scientific knowledge. These findings clearly indicate that the epistemic responsiveness of scholars is not limited to disciplinary concerns. Instead, by engaging in interdisciplinary encounters, both researchers and reviewers may be empowered to release themselves from narrow internal standards of their respective disciplines, which may in turn trigger more reflective consideration about the broader relevance of their work. For example, what may appear to lack rigor or excellence from a disciplinary point of view may still appear feasible, or even groundbreaking, from an interdisciplinary perspective (see Luukkonen 2012). My analyses of peer review deliberations suggest that such considerations can be triggered or constrained by the intersubjective context of evaluation. This finding points in the same direction as social psychological analyses (e.g. Shadish & Fuller 1994) and argumentation studies (e.g. Keith & Rehg 2008) of consensus formation in science, as well as constructivist analyses of sense-making in general (e.g. Eisner 1998). These analyses suggest that scientists, like everyone, can be regarded as having many subjective “selves” or mindsets, and which of them comes to the fore depends on the situation in which they find themselves. To occupy one of the many “trading zones” (Galison 1997) in which interdisciplinary accountabilities arise does not require the abandonment of one’s disciplinary “home”, or the loss of one’s primary identification as an expert in a given field. According to Nowotny and colleagues (2002, 177), “subtle accommodation processes may be at work, creating multiple and modified identities which are different”. The dissonance of knowledge in society, as well as the requirement for its management, has effects even at the level of the individual expert.

The concept of interdisciplinary accountability may be best understood from the perspective of “structuration”, as defined by Anthony Giddens (1984). According to this view, accountability relationships can be viewed as the structural properties of the governance system, which are maintained and reproduced by structured actions. At the same time, these actions—in this case, acts of account giving—have a transformative

capacity: Rule setters as well as rule followers are involved in information manipulation and strategic maneuvering (Yang 2011). My dissertation emphasizes the constitutive aspects of interdisciplinary account giving, including its use of interdisciplinary dialog as an *ad hoc* mechanism of epistemic accountability. Interdisciplinary accountability is thus embedded in epistemic relations in a state of flux, reflecting the contested and temporary nature of knowledge in general (cf. Yang 2011, 278).

5.3 An operational view of interdisciplinary accountability

A novelty of this dissertation is that it introduces the idea of interdisciplinary accountability as an operational concept in the evaluation of research proposals. While many authors in social epistemology, science policy, and related fields have identified the need for interdisciplinarity in the sense defined here (e.g. Campbell 1969; Frodeman & Mitcham 2004; Fuller 2000; Gulbenkian Commission 1996; Guston 2000; Weinberg 1962), there have been few pragmatic suggestions for interdisciplining knowledge. Fuller (2000), however, has explicitly analyzed and compared three possible strategies for transcending the internal standards of scientific quality: determining an explicit "finalization" of research fields (Schaefer 1983); including cross-disciplinary relevance as a criterion in science policy decisions (Weinberg 1962); and promoting forms of inquiry that are "epistemically fungible", i.e. likely to benefit several lines of research (Fuller 1993).

Unlike these strategies, my articulation of interdisciplinary accountability does not offer a clear vision of how to distribute research funds. What is does offer, however, is a relatively viable framework for making disciplinary communities routinely accountable to each other. As an alternative, horizontal form of epistemic accountability, interdisciplinarity challenges the taken-for-granted, hierarchical accountabilities assumed in disciplinary knowledge production—much in the same vein as feminist epistemologists, for example, who contend that "we have an epistemic responsibility […] to any person who takes our knowledge claims seriously" (Mauthner et al. 2002, 140; see also Grassvik 2011). Interdisciplinary accountability puts a premium on the *links, similarities, and parallel aspects* of research across disciplinary contexts, instead of on the distinctive culture of expertise within individual disciplines. It emphasizes the usability of knowledge resources in multiple intellectual settings, instead of resticting them to a single system of interpretation. The particular conceptions employed by researchers are thus considered less important than the epistemic accountability involved in making the conceptions as transparent as possible for the many communities who have a relationship to, or interest in, a given work. An interdisciplinary account should make sense to all those who work with related issues, not only to those from one's own intellectual field.

As evaluation necessarily operates within particular accounts, it is through the *forms and practices of account giving* that disciplines can be rendered more comparable. My findings illustrate that interdisciplinary accountability can, to some extent, be demonstrated and verified by using the basic form of a research proposal (e.g. Appendix 1). There is already a tendency towards more standardization of research evaluation

practices across countries and research fields (Strathern 2000a)—even in the social sciences and the humanities (see Lamont & White 2009), where differences between epistemic cultures are firmly institutionalized (Abbott 2001; Gulbenkian Commission 1996; Lamont 2009; Whitley 1984). Systematic reporting and checking of interdisciplinary accountabilities may extend this tendency into the epistemic core of science. Evaluation criteria, such as those used by the Academy of Finland—including scientific quality, innovativeness, and feasibility of the research plan, as well as competence and expertise of the research team (see Appendix 2)—could be conceived as nexuses of multiple disciplinary norms or open negotiation points, instead of as having sacrosanct, incommensurable properties that privilege a single disciplinary canon. The explicit inclusion of broader impact considerations within the review process, such as the criterion of societal impacts in the US National Science Foundation's process of merit review (see Holbrook & Frodeman 2011), is but one strategy to strengthen interdisciplinary accountability.

Interdisciplinary accountability, I suggest, can best be fostered by crediting research proposals on the basis of goal and process accountability for broad, multiple audiences. Funding agencies could instruct researchers about these dimensions and require them to actively search out and justify epistemic links to many relevant communities. A precise definition of accountability, and what kinds of research meet the definition or merit funding thereby, is beyond the scope of my dissertation. My findings, however, help defining the mechanism or process through which valid criteria can be found. I suggest using a particular social communicational device, interdisciplinary dialog, as a type of epistemic standard. What matters most is who is considered eligible to participate in evaluation, and on what basis. I concur with many others that we need to broaden our notion of "peer" to include others beside disciplinary specialists (Frederiksen et al. 2003; Fuller 2002; Holbrook 2010; Holbrook & Frodeman 2011; Klein 2008a). I also emphasize the need to cultivate the capacity for making context-sensitive judgments, including a readiness to rapidly shift perspectives or alter frames of reference due to communication with others (see also Russell 1983).

5.4 Interdisciplinary governance of knowledge

Disciplinary science is increasingly criticized for its reliance on "internal" sources of control, and thereby its lack of "external" accountability. Remedies have been sought from politicizing quality control in various ways, and transforming quality control from a professional mode of governance into a more democratic process. Indeed, in debates on the governance of science, many have argued for a democratized knowledge policy. This view is founded on a definition of knowledge that goes beyond the normative canon of disciplinary standards, and driven by the idea that epistemology is deeply political (e.g. Funtowicz & Ravetz 1990 & 1993; Fuller 2000 & 2002; Jasanoff 2004 & 2005). I share those concerns, but the sphere of politics taken into account here is restricted to the politics embedded in particular disciplinary epistemologies, and in the professional organization of intellectual production in general. As a mechanism of governance,

interdisciplinary quality control lies somewhere between professional and political accountability (as defined by Romzek & Dubnick 1987). Its source of control is neither internal nor external to disciplines; it emanates from a dialogue between experts with different internalized professional norms and standards, a dialogue that is based on an expectation of mutual responsiveness.

I acknowledge, however, that interdisciplining knowledge production is only one step in a more profound project of democratizing science. The instrumental or provisional stance of my argument can be contrasted with the more fundamental stance adopted by Fuller, for example. In his reflections on *The Governance of Science* (2000a), he states: "I assume that it is possible and desirable to construct a forum for 'knowledge policy' (understood in the broadest sense to cover both educational and research matters in all the academic disciplines) that would enable an entire society, regardless of expertise, to decide on which resources should be allocated to which projects, on the basis of which accountability structures" (Fuller 2000b, 527). As the author himself suspects, however, there may not be any locus of economic and political power that could realistically house such a forum. Further, I would also question the desirability of a centralized forum, and favor a more heterarchical mode of governance. While acknowledging that science has epistemic authority as delegated to it by the public (Jasanoff 2003), I maintain that the rules of epistemic authority, once established, cannot be changed without consulting the ruling bodies—otherwise we may easily end up with epistemic anarchy.

As a counterforce to disciplinary authority, the notion of interdisciplinary accountability highlights the critical functions of intellectual exchange between disciplines. Like scholarly critique and response in general, a more critical attitude between disciplines is likely to improve the reliability of knowledge (Fuller 1993; see also Campbell 1969). However, unlike the evolutionary theories of the development of knowledge, this dissertation highlights the role of interdisciplinary critique in an ecological or horizontal constitution of reliable knowledge. The notion of interdisciplinary accountability acknowledges that what is reliable in one context may not be so in another context, and what is needed is a knowledge culture characterized by lateral accountability, including monitoring and responsibility across disciplinary contexts. It measures worth by concepts such as "field rigor" (Frodeman 2010b). Indications of such a culture are currently visible in many fields of applied science, such as environmental research, where the "test" of reliable knowledge is ultimately the survival of our planet. Interdisciplinary accountability faces more challenges in pure academic fields, especially in the social sciences and the humanities (see Lamont 2009), or in fields currently characterized by a low degree of mutual dependence between scholars (Whitley 1984).

Mutual accountabilities across disciplinary boundaries are not exclusively about setting more restrictions, but also about multiplying assets. While interdisciplinarity is generally assumed to spawn innovation by cross-fertilizing or integrating disciplinary knowledges, competing criteria of evaluation are usually regarded as a barrier to this development (e.g. Messing 1996; Salter & Hearn 1996). The notion of interdisciplinary accountability, however, acknowledges that competing, yet coexisting principles of evaluation may be a source of innovation. As the cognitive challenge central to innovation and breakthrough is a "search during which you do not know what you are looking for but

will recognize it when you find it" (Stark 2009, 1), an ability to exploit existing knowledge while simultaneously allowing for unanticipated associations is essential (Abbott 2004; Stark 2009; Stefik & Stefik 2004). Both conceptual (e.g. Fuller 1988) and empirical (e.g. Carlile 2002; Stark 2009) analyses suggest that rival perspectives may effectively facilitate such reflexive cognition. Interdisciplinary breakthroughs, I argue, are most likely to happen when cognitive authority is distributed between mutually accountable disciplines.

Besides such instrumental values, however, interdisciplinary accountability can also be conceived as an end in itself. While deeper discussion on this aspect is outside the scope of this dissertation, mention of the idea is relevant here. Focusing on accountability as a voluntary aspiration to be answerable, arising from the ethics of scientific inquiry, may provide ballast for the increasing demands of auditability, governability, and other means of exercising control and scrutiny under the regime of New Public Management (see Cassin & Büttgen 2010). While this audit culture may enforce trustworthy behavior, it does not instill trust—it rather breeds suspicion (Bleiklie & Kogan 2007; Power 1997; Strathern 2000a & 2000b). In contrast, acknowledgement that the *values of science*—not only its outcomes—may be central to a good society, much in the same way as moral and aesthetic values, counters the claim that science is a continuation of politics by other means (Collins 2009 & 2012). Criteria for interdisciplinary accountability would thus come close to those for responsible research (e.g. McClintock et al. 2003). It is due to its epistemic virtuousness that interdisciplinary accountability may resonate with the values and the identity of academics. Instead of imposing direct rules on peer review deliberations, for example, or subjecting the process to various political (as opposed to scientific) goals, I suggest operating with the informal rules panelists themselves develop. The latter rules are crucial for the participants' faith in the peer review system, which, in turn, has a tremendous influence on how well the system works (Lamont 2009).

5.5 Limitations

The contribution of this dissertation is susceptible to several major limitations. While the limitations concerning the methodological and technical details of the original analyses are discussed in the articles, I concentrate here on the issues that concern the dissertation as a whole. These limitations have to do with my findings on interdisciplinary accountability in the evaluation of research proposals.

A first set limitations concerns the *construct validity* (Yin 2003, 35-36) of this dissertation. I have searched for answers to the theoretical puzzle of what constitutes interdisciplinary accountability, and how it can be demonstrated, validated, and strengthened in the evaluation of research proposals. To critically consider my contribution to the problematic, it is important to discuss to what extent I have studied valid evidence of it. The epistemic content of research proposals and the deliberation process of peer reviewers can offer some, but not extensive, evidence of the phenomenon. A more comprehensive picture would have been gained by including evidence from yet another level of proposal evaluation, the process by which the Research Councils of the

Academy of Finland make funding decisions on the basis of peer review statements. In addition, the classification of research proposals, on which the findings in Section 4.1 are based, is not necessarily the best approach to obtain information about the constituents of interdisciplinary accountability. A different scheme might arise from a more empirically driven approach, which could also go beyond the text of research proposals. Moreover, the classification scheme, as originally articulated in *Article II*, did not directly address the assumptions or forms of epistemic *accountability*; had it done so, it might have paid more attention to types of interdisciplinary interpenetration (as in Fuller 1993, Ch. 3) and degrees of intellectual control between fields (as in Whitley 1984, Ch. 5), for example. Some of the conceptual issues that follow from this discrepancy are, however, reflected upon in the beginning of Section 4.1. While such inconsistencies in my conception of interdisciplinarity may weaken the dissertation as a whole, they also reflect conceptual development and learning.

A second set of limitations concerns the *internal validity* (Yin 2003, 36) of the findings reported in Sections 4.2 and 4.3. Based on interview and documentary evidence of panel deliberations, the analysis suggest that certain conditions of those deliberations contribute to interdisciplinary accountability while some others hinder it. However, the studied panels differed also in other respects besides the given conditions. Various properties of the group, such as sex and age distribution, the number of participants, and the dynamics between personalities, for example, probably influence the deliberation rules, too (see Olbrech & Bornmann 2010). The effect of these factors on the emerging accountability relationships between panel members were not controlled in the analysis. Another set of important, but excluded factors pertains to the social motives of the participants, which may have an influence on the cognitive heuristics of groups (De Dreu & Carnevale 2003; Beersma & De Dreu 2003). Earlier research (Lamont 2009) and my interviews with panel members indicate that the social motives for participating in evaluations were very collegial and socially spirited as opposed to narrowly egotistic, which may be the most important driver for interdisciplinary accountability in panel deliberations. Given these concerns over internal validity, the findings offered in Sections 4.2 and 4.3 do not demonstrate causal processes; however, the findings do merit consideration as well-informed advice for the organizers of panel deliberations.

A third set of limitations concerns *external validity* (Yin 2003, 37), i.e. the generalization of my findings. I have studied the evaluation of research proposals in the context of a national funding agency in Finland, the Academy of Finland. This empirical focus brings about at least two restictions. First, Finland is a peculiar context of research funding, and does not correspond to the settings of many other countries. In countries like the US, for example, where the sheer volume of research activity is many times greater and the disciplinary structure of science much stronger, interdisciplinary accountability may confront more resistance and hostility, and thus require more sustained institutional changes to flourish (see also *Article IV*). In the UK, to take another example, the change towards an audit culture has been clearly more abrupt and wide-ranging than in Finland (Strathern 2000b; Whitley 2011, 366)—in the UK, models of interdisciplinary accountability may already be in use. Second, each funding agency has a unique strategy and its own profile of funding instruments. As the evaluation of Academy Projects

represents a case without any particular incentive for interdisciplinarity, stronger patterns of interdisciplinary accountability, at least at the rhetorical level, are likely to occur if interdisciplinarity is given priority by the funding agency.

The focus and design of my analysis of peer review deliberations set another limit to the external validity of findings reported in Sections 4.2 and 4.3, in particular. Those findings were based on a comparison of the deliberation processes of different peer review panels, all of which considered proposals in social sciences and humanities, and/or in environmental sciences. Deliberations in other fields may follow quite different rules, and somewhat different factors may come up. Some differences between fields were indeed recognized between the social sciences and humanities, on the one hand, and environmental sciences, on the other (see also *Article III*). As implied by Whitley's analysis in *The Intellectual and Social Organization of the Sciences* (1984; see also Whitley et al. 2010), it may be that the whole concept of interdisciplinary accountability is more relevant in natural sciences and in applied fields that in the humanities and the social sciences.

A final set of limitations concerns the *reliability* (Yin 2003, 37-39) of this dissertation. A specific reliability test was conducted as a part of the analysis on which the findings in Section 4.1 are based (see *Article II*). The test was designed to measure the reliability of the judgments made by a subjective analyzer, i.e. myself, in categorizing research proposals. While the inter-rater reliability test showed no significant level of correlation between my results and those of another classifier of our team, this does not necessarily imply that the suggested (sub-)categories of interdisciplinary accountability as such are unsound, especially because we noticed that a discussion between the two classifiers quickly led to mutual understanding of proposals. What it does imply, however, is that identifying and categorizing interdisciplinary accountabilities is necessarily laborious (see Appendix 6) and cannot realistically be conducted without having expertise of the scientific topic itself. This is a further reason to emphasize the role of researchers and reviewers themselves in reporting and checking accountabilities across disciplinary boundaries.

6 Conclusions

6.1 Implications for research evaluation

This dissertation has investigated epistemic accountabilities across disciplines by drawing on the literature on interdisciplinarity as well as on empirical studies conducted in a national research funder in Finland. On this basis, it has suggested concepts and practices for more interdisciplinary policies of research evaluation. The findings provide a counterpoint to the prevailing ideas of academic quality control by reconsidering the normative infrastructure on which such ideas rest. The major issue is to open scholarly knowledge production to scrutiny from the outside—not only from the public, but also from the other fields of research. An interdisciplinary model of evaluation would redistribute epistemic authority according to lines of lateral accountability: While the disciplinary model holds knowledge accountable to the given disciplinary community alone, the interdisciplinary model makes disciplinary norms open to negotiation, change, and local modification. The key is thus not to give up on disciplinary regimes of knowledge, but to provide communicative forms and channels for renegotiating their meaning and generating new epistemic standards (as in Fuller 1993).

As a concept in research evaluation, interdisciplinarity is *movement* in epistemic norms. The notion of interdisciplinary accountability shifts the analytical focus from the normative structure of science to questions of *process*: Where do interdisciplinary norms come from? The pragmatic challenge it poses for research evaluation is how to make legitimate appraisals on the basis of moving parameters. For this purpose, the dissertation has articulated a framework for conceptualizing interdisciplinary accountability in research proposals, and considered options to facilitate interdisciplinary accountability through peer review.

The rationale for interdisciplining evaluation is to increase the scientific community's capacity to assess the robustness, reliability and relevance of academic knowledge beyond the self-contained bubble of any one specialty. An interdisciplinary policy of research evaluation would be committed to managing the dissonance of knowledge and overcoming the myopia of disciplinary specialism within the normative framework of science, i.e. without imposing commercial, national, or political interests on science, though aware that claims about knowledge are never completely free of politics. The goal of interdisciplining evaluation is to transform the central mechanism of the governance of knowledge from being the handmaiden of the disciplines to serving science—and ultimately, the knowledge society as a whole. Many pervasive problems in evaluating interdisciplinary research would be remedied at the same time.

Such a move does not imply only that disciplinary research would be subjected to broader scrutiny, but also that interdisciplinary research would be held in check by its disciplinary antecedents. This view builds on the existing knowledge of evaluating interdisciplinary research (Boix Mansilla 2006; Klein 2006 & 2008a), but is distinguished from most studies by addressing the interdependencies that occur between disciplines in all forms of research, instead of focusing on the characteristics peculiar to interdisciplinary

research. This even-handed treatment of all research is advantageous to research evaluation in several respects. First, it removes the problem of how to define and identify interdisciplinary research as opposed to disciplinary research, which has proved both conceptually and pragmatically to be an enormous task (see Appendix 6 and *Article II*). Second, it avoids the epistemically undesirable consequence that interdisciplinarity becomes either used as a residual category of evaluation, to include what reviewers regard as unorthodox research, or the insurmountable ideal of fulfilling the criteria of several disciplines at the same time. Third, it goes against the assertion that interdisciplinarity is only a fad, without any lasting influence on the epistemic core of knowledge production.

Yet those who execute research evaluations in practice may need more clarity in how to go about evaluating interdisciplinary proposals in the here and now; while they may agree that no clear division exists between disciplinary and interdisciplinary research, they also witness a gap between the current practices of evaluation and those proposals that do not fit in. They may find a selective use of the concepts and practices suggested by this dissertation useful. Specifically, they could direct interdisciplinary researchers to lay out their proposals according to the framework offered in Section 4.1, and design some peer review panels with a view to facilitate interdisciplinary accountability. This strategy may then be broadened to encourage other researchers to actively consider their projects across and among other relevant disciplines.

I suggest the following interdisciplinary strategy for evaluating research proposals:

1. Encourage both researchers and reviewers to consider their epistemic accountabilities to multiple disciplinary communities in terms of both the desired goals of research and the process of accomplishing those goals. Instead of drawing boundaries between fields, encourage researchers and reviewers to search for links and parallels across fields of knowledge. Emphasis should be put on the relevance, usability, and reliability of knowledge in multiple intellectual settings.
2. Instruct researchers to lay out their proposals with an eye toward convincing a wider academic audience than only experts in their own field. Proposals need to be as transparent as possible for the many communities who have a relationship to, or interest in, the given work. They should explicitly address the relevant epistemic stakeholders, the functions and beneficiaries of the proposed research, and the process by which knowledge and experts from different fields can be utilized.
3. In designing peer review panels, ensure both multidisciplinary competence and interdisciplinary dialog. Mix experts from different, but not disparate fields. Select generalist panel members who possess a broad knowledge beyond any one academic field.

6.2 Future research needs

This dissertation started from a pragmatic need to reconsider the concepts and practices of research evaluation in the face of interdisciplinary desires and pressures. It has opened up

the problematic of interdisciplinary accountability in the evaluation of research proposals, and provided preliminary ideas and observations on the topic. Given the selected focus and limitations addressed in Section 5.5, further research is needed to fill in the gaps. In particular, two areas for future research arise directly from the need to develop more diversified and powerful tools for increasing interdisciplinary accountability in research evaluation.

First, the practices of interdisciplinary accountability in peer review panels need to be investigated more systematically. The various integrative or consensual strategies of peer negotiations found in this research illustrate that differences do exist, but larger samples and a broader range of panels will have to be considered before generalizations about trends and causal processes can be made. How can interdisciplinary accountability be best facilitated in different fields of science? What is the optimal degree of overlap in reviewers' expertise to ensure that good interdisciplinary dynamics will arise in various evaluative settings? What formal criteria or technologies of evaluation can be used to facilitate interdisciplinary accountability, and what is an appropriate balance between structured and unstructured procedures? Besides more empirical evidence of these issues, more systematic use of the methodological apparatuses of social psychology (e.g. Olbrecht & Bornmann 2011; Thorngate et al. 2009; Shadish & Fuller 1994) and negotiation research (e.g. Beersma & De Dreu 2003) is needed.

Second, other empirical and policy contexts of the evaluation of research need to be investigated. The present analysis helps to define and facilitate interdisciplinary accountability at the early stage of research evaluation, i.e. in the evaluation of research proposals; it could be complemented by investigating accountabilities in the evaluation of the products of research and the applicants for academic posts. Existing analyses of peer review deliberations on article manuscripts (e.g. Hirschauer 2010) and book submissions (Powell 1985) imply somewhat different patterns of accountability, but little is known about interdisciplinary considerations in these settings. Variations across national settings could also be investigated.

This dissertation has also inspired new questions about interdisciplinary accountability. A problem area that deserves more attention concerns interdisciplinary accountability not only as an analytical concept in research evaluation, but as a desirable state: What does "being accountable" mean in various interdisciplinary contexts? To answer this question, both empirical and philosophical investigations are needed. Interview studies on what defines good interdisciplinary research do exist (e.g. Boix Mansilla 2006), but a more concrete picture could be gained by, for example, analyzing the properties of more and less meritorous interdisciplinary proposals vis-à-vis disciplinary ones, as well as reasons offered for their intellectual merit. What kind of interdisciplinary criteria or heuristics do experts use to negoatiate, for instance, what research goals are of top priority? More intensive dialog between scholars on interdisciplinarity and scholars on distributed cognition would yield new insights on these issues (e.g Derry et al. 2005). Philosophy of science and argumentation theory are also needed to complement empirical research.

Another new area of interest for future researchers is the role of interdisciplinary accountability in relation to other accountabilities that exist in knowledge production and evaluation. Disciplinary expectations are not likely to disappear, and new demands for

productivity and impact are being expressed by various public and private actors outside the realm of science. What is the relationship between interdisciplinary accountability and other forms of accountability, and what are the effects of multiple or contradictory expectations? In contrast to the claims of some theorists of interdisciplinarity (Barry et al. 2008; Strathern 2004), interdisciplinary accountability may not be an unambiguous marker of social accountability. It is argued, for instance, that interdisciplinary research would be less inclined to contribute to pragmatic decision-making if the communication between disciplines focuses on epistemic robustness at the expense of social robustness (Sarewitz 2010; Stirling 2010; see O'Rourke, draft). Multiple accountabilities could also be investigated from the normative perspective of research evaluation, which means adopting a more complex view on the problematic addressed in this dissertation. What might a research evaluation look like that strikes the right balance of accountability to the various kinds of stakeholders in knowledge production? Obviously, trying to be accountable to everyone might ultimately end up in accountability to no one (Dubnic & Frederickson 2011; Strathern 2000a). Relationships between professional, epistemic, political, administrative, and other values in research evaluation need to be explored further.

Future research should also investigate the relationship between accountability and evaluation in general, which is conceptually and empirically more complex than I have made it to be here (see e.g. Cassin & Büttgen 2010). I have implied in this dissertation that evaluation is a means to impose control through accountability. On the other side, accountability regarded as a form of epistemic responsiveness or responsibility is a virtue very difficult to call into question. I have assumed that accountability serves the purposes of evaluation, but it could also be the other way around. Perhaps we should evaluate and credit research more on the basis of the ethics of responsibility; interdisciplinary accountability would be a constitutive element of such a framework. More light on this issue could be shed from the ethics of science and from virtue epistemology (Greco & Turri 2011; Zagzebski 1996). A very different yet potential source of insight is the professional discourse on evaluation, which addresses evaluation as assisted sense-making that can support social betterment (e.g. Mark et al. 2000; Schwandt 2002). If accountability is a virtue with which research is increasingly associated, we need to look at how evaluation practitioners go about determining it. What is distinctive in their discourse is a circumspect attitude, even resistance, to the further professionalization of their practice, and the emphasis they put on their role as reflective facilitators.

References

Abbott A (2001) *Chaos of Disciplines*. Chicago & London: The University of Chicago Press.

Abbott A (2004) *Methods of Discovery: Heuristics for the Social Sciences*. Series: Alexander JC (Ed), Contemporary Societies. New York & London: W.W. Norton & Company.

Academy of Finland (2012) *Research knows no boundaries: Annual report 2011*. Helsinki: Academy of Finland. Available at: http://www.aka.fi/Tiedostot/Tiedostot/Akatemia_vk_2011_taitto_EN%20%282%29.pdf

Alkin MC (1972) Accountability defined. *Evaluation Comment: The Journal of Educational Evaluation* 3: 1–5.

Alkin MC (2004) *Evaluation Roots: Tracing Theorists' Views and Influence*. Los Angeles & London: SAGE.

Balsamo A and Mitcham C (2010) Interdisciplinarity in ethics and the ethics of interdisciplinarity. In: Frodeman R, Klein JT and Mitcham C (Eds), *The Oxford Handbook of Interdisciplinarity*. Oxford: Oxford University Press. Pp. 259–272.

Barry A, Born G and Weszkalnys G (2008) Logics of interdisciplinarity. *Economy and Society* 37: 20–49.

Beersma B and De Dreu CKW (2003) Social motives in integrative negotiation: The mediating influence of procedural fairness. *Social Justice Research* 16(3): 217–239.

Bergmann M, Brohmann B, Hoffmann E, Loibl MC, Rehaag R, Schramm E and Voß J-P (2005) *Quality Criteria of Transdisciplinary Research: A Guide for the Formative Evaluation of Research Projects*. Frankfurt am Main: Institute for Social-Ecological Research (ISOE).

Bleiklie I and Kogan M (2007) Organization and governance of universities. *Higher Education Policy* 20: 477–493.

Boden MA (1999) What is interdisciplinarity? In: Cunningham R (Ed), *Interdisciplinarity and the Organization of Knowledge in Europe*. A Conference Organised by the Academia Europaea. Cambridge, September 24–26, 1997. Luxembourg: Office for Official Publication of the European Communities. Pp. 13–24.

Boix Mansilla V (2006) Assessing expert interdisciplinary work at the frontier: An empirical exploration. *Research Evaluation* 15(1): 17–29.

Boix Mansilla V (2010) Learning to synthesize: The development of interdisciplinary understanding. In: Frodeman R, Klein JT and Mitcham C (Eds), *The Oxford Handbook of Interdisciplinarity*. Oxford: Oxford University Press. Pp. 288-306.

Boix Mansilla V, Feller I and Gardner H (2006) Quality assessment in interdisciplinary research and education. *Research Evaluation* 15(1): 69–74.

Bowker GC and Star SL (1999) *Sorting Things Out: Classification and Its Consequences*. Cambridge, MA: MIT Press.

Braun D and Merrien F-X (Eds) (1999) *Towards a New Model of Governance for Universities? A Comparative View*. London: Jessica Kingsley Publishers.

Bruce A, Lyall C, Tait J and Williams R (2004) Interdisciplinary integration in Europe: The case of the Fifth Framework programme. *Futures* 36: 457–470.

Bruun H, Hukkinen J, Huutoniemi K and Klein JT (2005) Promoting interdisciplinary research: The case of the Academy of Finland. *Publications of the Academy of Finland* 8/05. Helsinki: The Academy of Finland.

Bruun H, Langlais R and Janasik N (2005) Knowledge networking: A conceptual framework and typology. *VEST* 18(3-4): 73–104.

Bruun H and Toppinen A (2004) Knowledge in science and innovation: A review of three discources on the institutional and cognitive foundations of knowledge production. *Issues of Integrative studies* 22: 1–51.

Burawoy M (2005) For public sociology. *American Sociological Review* 70: 4–28.

Cambrosio A and Keating P (1983) The disciplinary stake: The case of chronobiology. *Social Studies of Science* 13: 323–353.

Campbell DT (1969) Ethnocentrism of disciplines and the fish-scale model of omniscience. In: Sherif M and Sherif CW (Eds), *Interdisciplinary Relationships in the Social Sciences*. Chicago: Aldine. Pp. 328–48.

Carlile PR (2002) A pragmatic view of knowledge and boundaries: Boundary objects in new product development. *Organization Science* 13(4): 442–455.

Carlile PR (2004) Transferring, translating, and transforming: An integrative framework for managing knowledge across boundaries. *Organization Science* 15(5): 555–568.

Cassin B and Büttgen P (2010) The performative without condition: A university sans appel. *Radical Philosophy* 162: 31–37.

Chubin DE and Hackett EJ (1990) *Peerless Science: Peer Review and U.S. Science Policy*. Albany, NY: State University of New York Press.

Cohen WM and Levinthan DA (1990) Absorptive capacity: A new perspective on learning and innovation. *Administrative Science Quarterly* 35(1): 128–152.

Collins HM (2009) We cannot live by skepticism alone. *Nature* 458: 30–31 (5 March 2009)

Collins HM (2012) Elective modernism. Available at: http://www.cf.ac.uk/socsi/contactsandpeople/harrycollins/expertise-project/draftpapers/index.html

Collins HM and Evans R (2007) *Rethinking Expertise*. Chicago, IL: University of Chicago Press.

Considine M (2002) The end of the line? Accountable governance in the age of networks, partnerships, and joined-up services. *Governance: An International Journal of Policy, Administration, and Institutions* 15(1): 21–40.

Cunningham R (Ed) (1997) *Interdisciplinarity and the Organization of Knowledge in Europe*. A Conference Organised by the Academia Europaea. Cambridge, September 24–26, 1997. Luxembourg: Office for Official Publication of the European Communities.

De Dreu CKW and Carnevale PJ (2003) Motivational bases of information processing and strategy in conflict and negotiation. *Advances in Experimental Social Psychology* 35: 235–291.

Derry SJ, Schunn CD and Gernsbacher MA (Eds) (2005) *Interdisciplinary Collaboration: An Emerging Cognitive Science*. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Doucet A and Mauthner N (2002) Knowing responsibly: Ethics, feminist epistemologies and methodologies. In: Mauthner M, Birch M, Jessop J and Miller T (Eds), *Ethics in Qualitative Research*. London: Sage. Pp. 123–145.

Dreyfus HL and Dreyfus SE (2005) Expertise in real world contexts. *Organization Studies* 26, 779–792.

Dubnick MJ and Frederickson HG (2011) Introduction: The promises of accountability research. In: Dubnick MJ and Frederickson HG (Eds), *Accountable Governance. Problems and Promises*. Armonk, NY: M.E.Sharpe. Pp. xiii–xxxii.

Eisenhart M (2002) The paradox of peer review: Admitting too much or allowing too little? *Research in Science Education* 32: 241–255.

Eisner EW (1998) *The Enlightened Eye. Qualitative Inquiry and the Enhancement of Educational Practice*. Upper Saddle River, New Jersey & Columbus, Ohio: Prentice Hall.

Elzinga A (1993) Science as the continuation of politics by other means. In: Brante T, Fuller S and Lynch W (Eds), *Controversial Science: From Content to Contention.* Albany: State University of New York. Pp. 127–152.

Espeland W and Stevens M (1998) Commensuration as a social process. *Annual Review of Sociology* 24: 313–343.

European Union Research Advisory Board (EURAB) (2004) *Interdisciplinarity in Research*. Available at: http://ec.europa.eu/research/eurab/pdf/eurab_04_009_ interdisciplinarity_research_final.pdf

Ferlie E, Musselin C and Andresani G (2009) The 'steering' of higher education systems: A public management perspective. In: Paradeise C, Reale E, Bleiklie I and Ferlie E (Eds), *University Governance*: *Western European Comparative Perspectives.* Dordrecht: Springer. Pp. 1–20.

Frederiksen LF, Hansson F and Wenneberg SB (2003) The agora and the role of research evaluation. *Evaluation* 9(2): 149–172.

Frodeman R (2010a) Introduction. In: Frodeman R, Klein JT and Mitcham C (Eds), *The Oxford Handbook of Interdisciplinarity*. Oxford: Oxford University Press. Pp. xxix–xxxix.

Frodeman R (2010b) Experiments in field philosophy. *Opinionator. Exclusive Online Commentary From The Times.* Available at: http://opinionator.blogs.nytimes.com/ 2010/11/23/experiments-in-field-philosophy/

Frodeman R (2011) Interdisciplinary research and academic sustainability: Managing knowledge in an age of accountability. *Environmental Conservation* 38(2): 105–112.

Frodeman R and Mitcham C (2004) New dimensions in the philosophy of science: Toward a philosophy of science policy. *Philosophy Today* 48(5): 3–14.

Frodeman R and Mitcham C (2007) New directions in interdisciplinarity: Broad, deep, and critical. *Bulletin of Science, Technology & Society* 27(6): 506–14.

Fuller S (1988) *Social Epistemology*. Bloomington: Indiana University Press.

Fuller S (1993) *Philosophy, Rhetorics, and the End of Knowledge: The Coming of Science and Technology Studies*. Madison, WI: University of Wisconsin Press.

Fuller S (2000a) *The Governance of Science: Ideology and the Future of the Open Society.* Philadelphia: Open University Press.

Fuller S (2000b) Increasing science's governability: Response to Hans Radder. *Science, Technology, & Human Values* 25(4): 527–534.

Fuller S (2002) *Knowledge Management Foundations.* Boston, MA: Butterworth-Heinemann.

Fuller S (2010) Deviant interdisciplinarity. In: Frodeman R, Klein JT and Mitcham C (Eds), *The Oxford Handbook of Interdisciplinarity.* Oxford: Oxford University Press. Pp. 50–64.

Funtowicz SO and Ravetz JR (1990) *Uncertainty and Quality in Science for Policy*. Dordrecht: Kluwer.

Funtowicz SO and Ravetz JR (1993) Science for the post-normal age. *Futures* 25(7), 739–55.

Galison P (1997) *Image and Logic: A Material Culture of Microphysics.* Chicago: University of Chicago Press.

Garfinkel H (1967) *Studies in Ethnomethodology*. Englewood Cliffs, NJ: Prentice Hall.

Gibbons M, Dowling PJ, Mirdal G and Petterson RF (2004) *International Evaluation of the Academy of Finland*. Helsinki, Ministry of Education, Department of Education and Science Policy, 2004:16. Available at: http://www.minedu.fi/export/sites/default/ OPM/Julkaisut/2004/liitteet/opm_232_opm16.pdf?lang=fi

Gibbons M, Limoges C, Nowotny H, Schwartzman S, Scott P and Trow M (1994) *The New Production of Knowledge: The dDynamics of Science and Research in Contemporary Societies*. London: Sage.

Gibbons M and Nowotny H (2001) The potential of transdisciplinarity. In: Klein JT, Grossenbacher-Mansuy W, Scholz RW and Welti M (Eds), *Transdisciplinarity: Joint Problem Solving among Science, Technology, and Society*. Basel: Birkhäuser. Pp. 67–80.

Giddens A (1984) *The Constitution of Society: Outline of the Theory of Structuration*. Berkeley: University of California Press.

Gieryn T (1999) *Cultural Boundaries of Science*. Chicago: University of Chicago Press.

Goldman A (2010) Social Epistemology. In: Zalta EN (Ed), *The Stanford Encyclopedia of Philosophy* (Summer 2010 Edition). Available at: http://plato.stanford.edu/archives/sum2010/entries/epistemology-social/

Grassvik H (Ed) (2011) *Feminist Epistemology and Philosophy of Science: Power in Knowledge*. Dordrecht: Springer.

Greco J and Turri J (2011) Virtue Epistemology. In: Zalta E (Ed), *The Stanford Encyclopedia of Philosophy (Winter 2011 Edition)*. Available at: http://plato.stanford.edu/archives/win2011/entries/epistemology-virtue/

Grigg L (1999) *Cross-Disciplinary Research. A Discussion Paper*. Commissioned Report No. 61. Canberra: Australian Research Council.

Guba E and Lincoln Y (1989) *Fourth Generation Evaluation*. Newbury Park, CA: Sage.

Gulbenkian Commission (1996). *Open the Social Sciences*. Report of the Gulbenkian Commission on the Restructuring of the Social Sciences. Stanford, CA: Stanford University Press.

Gulbrandsen JM (2000) *Research Quality and Organisational Factors: An Investigation of the Relationship*. Doctoral dissertation. Department of Industrial Economics and Technology Management, Norwegian University of Science and Technology, Trondheim.

Guston D (2000) *Between Politics and Science: Assuring the Integrity and Productivity of Research*. New York: Cambridge University Press.

Hackett EJ and Chubin DE (2003) *Peer Review for the 21st Century: Applications to Education Research*. Prepared for a National Research Council Workshop, Washington, D.C.

Hirschauer S (2010) Editorial judgments: A praxeology of 'voting' in peer review. *Social Studies of Science* 40: 71–103.

Holbrook JB (2010) Peer review. In: Frodeman R, Klein JT and Mitcham C (Eds), *The Oxford Handbook of Interdisciplinarity*. Oxford: Oxford University Press. Pp. 321–331.

Holbrook JB and Frodeman R (2011): Peer review and the *ex ante* assessment of societal impacts. *Research Evaluation* 20(3): 239–246.

Huutoniemi K (2004) Realismi ja konstruktivismi tieteidenvälisessä ympäristötutkimuksessa—yhteismitattomuus, yhteensovittaminen vai dialogi? *Alue ja Ympäristö* 33(1): 3–15. [In Finnish]

Jacobs JA and Frickel S (2009) Interdisciplinarity: A critical assessment. *Annual Review of Sociology* 35: 43–65.

Janis I (1972) *Groupthink*. Boston, MA: Houghton-Mifflin.

Jasanoff S (2003) (No?) Accounting for expertise. *Science and Public Policy* 30(3): 157–162.

Jasanoff S (2004) *States of Knowledge: The Co-Production of Science and the Social Order*. London: Routledge.

Jasanoff S (2005) *Designs on Nature: Science and Democracy in Europe and the United States*. Princeton, NJ: Princeton University Press.

Jefftrey P (2003) Smoothing the waters: Observations on the process of cross-disciplinary research collaboration. *Social Studies of Science* 33(4): 539–562.

Kastenhofer K, Omann I, Stagl S and Steininger K (2002) Science policy for transdisciplinary research. In: Hirsch Hadorn G (Ed), *Unity of Knowledge in Transdisciplinary Research for Sustainability*. Encyclopedia of Life Support Systems (EOLSS). Oxford: Eolss Publishers. Available at: http://www.eolss.net

Kearns KP (2011) Accountability in the nonprofit sector: Abandoning the one-size-fits-all approach. In: Dubnick MJ and Frederickson HG (Eds), *Accountable Governance: Problems and Promises*. Armonk, NY: M.E.Sharpe. Pp. 197–210.

Keith W and Rehg W (2008) Argumentation in science: The cross-fertilization of argumentation theory and science studies. In: Hackett EJ, Amsterdamska O, Lynch M and Wajcman J (Eds), *The Handbook of Science and Technology Studies*. Third edition. Cambridge, MA: The MIT Press. Pp. 211–239.

Kelly JS (1996) Wide and narrow interdisciplinarity. *The Journal of Education* 45, 95–113.

Klein JT (1990) *Interdisciplinarity: History, theory, and practice*. Detroit, MI: Wayne State University Press.

Klein JT (1996) *Crossing Boundaries: Knowledge, Disciplinarities, and Interdisciplinarities*. Charlottesville, VA: University Press of Virginia.

Klein JT (2005) *Humanities, Culture, and Interdisciplinarity: The Changing American Academy*. Albany, NY: SUNY Press.

Klein JT (2006) Afterword: The emergent literature on interdisciplinary and transdisciplinary research evaluation. *Research Evaluation* 15(1): 75–80.

Klein JT (2008a) Evaluation of interdisciplinary and transdisciplinary research: A literature review. *American Journal of Preventive Medicine* 35, S116–S123.

Klein JT (2008b) The rhetoric of interdisciplinarity: Boundary work in the construction of new knowledge. In: Lunsford A and Aune J (Eds), *Handbook on Rhetoric*. Los Angeles & London: SAGE. Pp. 265–283.

Klein JT (2010) A taxonomy of interdisciplinarity. In: Frodeman R, Klein JT and Mitcham C (Eds), *The Oxford Handbook of Interdisciplinarity*. Oxford: Oxford University Press. Pp. 15–30.

Knorr Cetina K (1999) *Epistemic Cultures: How the Sciences Make Knowledge*. Cambridge, MA: Harvard University Press.

Krohn W (2010) Interdisciplinary cases and disciplinary knowledge. In: Frodeman R, Klein JT and Mitcham C (Eds), *The Oxford Handbook of Interdisciplinarity*. Oxford: Oxford University Press. Pp. 31–49.

Kunseler E (2007). Towards a new paradigm of science in scientific policy advising. Available at: http://www.nusap.net/downloads/KunselerEssay2007.pdf

Lamont M (2009) *How Professors Think: Inside the Curious World of Academic Judgment*. Cambridge, MA: Harvard University Press.

Lamont M and White P (2009) *Workshop on Interdisciplinary Standards for Systematic Qualitative Research*. Washington: National Science Foundation. Available at: http:// www.nsf.gov/sbe/ses/soc/ISSQR_workshop_rpt.pdf

Langfeldt L (2004) Expert panels evaluating research: Decision-making and sources of bias. *Research Evaluation* 13(1): 52–62.

Langfeldt L (2006) The policy challenges of peer review: Managing bias, conflict of interests and interdisciplinary assessments. *Research Evaluation* 15(1): 31–41.

Lattuca LR (2001) *Creating Interdisciplinarity: Interdisciplinary Research and Teaching among College and University Faculty*. Nashville, TN: Vanderbilt University Press.

Laudel G (2006) Conclave in the Tower of Babel: How peers review interdisciplinary research proposals. *Research Evaluation* 15(1): 57–68.

Levins R (1968) *Evolution in Changing Environments*. Princeton, NJ: Princeton University Press.

Lindberg SI (2009) *Accountability: The Core Concept and Its Subtypes*. Working Paper No. 1. Overseas Development Institute, London: African Power and Politics Research Program.

Longino HE (2002) *The Fate of Knowledge*. Princeton, NJ: Princeton University Press.

Luukkonen T (2002) Research evaluation in Europe: The state of the art. *Research Evaluation* 11(2): 81–84.

Luukkonen T (2012) Conservatism and risk-taking in peer review: Emerging ERC practices. *Research Evaluation* 21: 48–60.

Maasen S, Lengwiler M and Guggenheim M (2006) Practices of transdisciplinary research: Close(r) encounters of science and society. *Science and Public Policy* 33(6): 394–398.

Maasen S, Lieven O (2006) Transdisciplinarity: A new mode of governing science? *Science and Public Policy* 33(6): 399–410.

Mallard G, Lamont M and Guetzkow J (2009) Fairness as appropriateness: Negotiating epistemological differences in peer review. *Science, Technology and Human Values* 34(5): 573–606.

Mark M, Henry G and Julnes G (2000). *Evaluation: An Integrated Framework for Understanding, Guiding, and Improving Public and Nonprofit Policies and Programs.* San Francisco, CA: Jossey-Bass.

Marsh HW, Jayasinghe UW and Bond NW (2008) Improving the peer review process for grant applications. *American Psychologist* 63(3): 160–168.

Mauthner M, Birch M, Jessop J, Miller T (Eds) (2002) *Ethics in Qualitative Research*. London: Sage.

McClintock D, Ison R and Armson R (2003) Metaphors for reflecting on research practice: Researching *with* people. *Journal of Environmental Planning and Management* 46(5): 715–731.

Merton RK [1942] (1973) The normative structure of science. In: Storer NW (Ed), *The Sociology of Science*: *Theoretical and Empirical Investigations.* Chicago, IL: The University of Chicago Press.

Messing K (1996) Easier said than done. Biologists, economists, and sociologists collaborate to study health effects of the sexual division of labour. In: Salter L and Hearn A (Eds), *Outside the Lines: Issues in Interdisciplinary Research.* Montreal, QC: McGill-Queen's University Press. Pp. 95–102.

Miller R (1982) Varieties of interdisciplinary approaches in the social sciences. *Issues in Integrative Studies* 1: 1–37.

Mollinga PP (2008) The rational organization of dissent. Boundary concepts, boundary objects and boundary settings in the interdisciplinary study of natural resources management. *ZEF Working Paper Series* 33. Center for Development Research, University of Bonn.

Morin E (2008) *On Complexity*. Cresskill, NJ: Hampton Press.

National Academy of Sciences (2005) *Facilitating interdisciplinary research.* Washington, DC: National Academies Press.

Nikitina S (2005) Interdisciplinary cognition. *Cognition and Instruction* 23(3): 389–425.

Nowotny H (2003) Democratising expertise and socially robust knowledge. *Science and Public Policy* 30(3): 151–156.

Nowotny H (2006) Real science is excellent science: How to interpret post-academic science, Mode 2, and the ERC. *Journal of Science Communication* 5(4): 1–3.

Nowotny H, Scott P, Gibbons M (2002) *Re-Thinking Science: Knowledge and the Public in an Age of Uncertainty*. Cambridge, UK: Polity Press.

Olbrecht M and Bornmann L (2010) Panel peer review of grant applications: What do we know from research in social psychology on judgment and decision-making in groups? *Research Evaluation* 19(4): 293–304.

Organisation for Economic Cooperation Development (OECD) (1982). *The University and the Community. The Problems of Changing Relationships.* Paris: OECD/CERI.

O'Rourke M (draft) Enhancing communication in interdisciplinary research: A means to what end? Available at: http://www.webpages.uidaho.edu/~morourke/Toolbox/ECCDR-Book/Proposals/19.pdf

Palmer CL (2001) *Work at the Boundaries of Science: Information and the Interdisciplinary Research Process.* Dordrecht: Kluwer Academic Publishers.

Pielke Jr RA and Byerly Jr R (1998) Beyond basic and applied. *Physics Today* 51(2): 42–46.

Pohl C and Hirsch Hadorn G (2007) *Principles for Designing Transdisciplinary Research: Proposed by the Swiss Academies of Arts and Sciences.* München: oekom Verlag.

Polanyi M (1962) The republic of science: Its political and economic theory. *Minerva* 1(1): 54–73.

Porter AL and Rossini FA (1984) Interdisciplinary research redefined: Multi-skill, problem-focussed research in the STRAP framework. *R&D Management* 14: 105–111.

Porter AL and Rossini FA (1985) Peer review of interdisciplinary research proposals. *Science, Technology and Human Values* 10(3): 33–38.

Porter AL, Roessner JD, Cohen AS and Perreault M (2006) Interdisciplinary research: Meaning, metrics and nurture. *Research Evaluation* 15(3): 187–195.

Powell WW (1985) *Getting into Print. The Decision-Making Process in Scholarly Publishing*. Chicago, IL: The University of Chicago Press.

Power M (1997) *The Audit Society: Rituals of Verification.* Oxford: Oxford University Press.

Rafols I (2007) Strategies for knowledge acquisition in bionanotechnology: Why are interdisciplinary practices less widespread than expected? *Innovation* 20(4): 395–412.

Repko AF (2008) *Defining Interdisciplinary Studies.* London: Sage.

Research Evaluation (2006) Volume 15(1): Special issue on interdisciplinary research assessment.

Rogers Y, Scaife M and Rizzo A (2005) Interdisciplinarity: An emergent of engineered process? In: Derry SJ, Schunn CD and Gernsbacher MA (Eds), *Interdisciplinary Collaboration: An Emerging Cognitive Science.* Mahwah, NJ: Lawrence Erlbaum Associates. Pp. 265–285.

Romzek BS (2011) The tangled web of accountability in contracting networks: The case of welfare reform. In: Dubnick MJ and Frederickson HG (Eds), *Accountable Governance: Problems and Promises.* Armonk, NY: M.E.Sharpe. Pp. 22–41.

Romzek BS and Dubnick MJ (1987) Accountability in the public sector: Lessons from the Challenger Tragedy. *Public Administration Review* 47: 227–238.

Russell MG (1983) Peer review in interdisciplinary research: Flexibility and responsiveness. In: Epton SR, Payne RL and Pearson AW (Eds), *Managing Interdisciplinary Research.* New York, NY: John Wiley & Sons. Pp. 184–202.

Salter L and Hearn A (1996) Interdisciplinarity. In: L. Salter and A. Hearn (Eds), *Outside the Lines: Issues in Interdisciplinary Research.* Montreal: McGill-Queen's University Press.

Sarewitz D (2000) Science and environmental policy: An excess of objectivity. In: Frodeman, R (Ed), *Earth Matters: The Earth Sciences, Philosophy, and the Claims of Community*. Upper Saddle River, NJ: Prentice Hall. Pp. 79–98.

Sarewitz D (2010) Against holism. In: Frodeman R, Klein JT and Mitcham C (Eds), *The Oxford Handbook of Interdisciplinarity*. Oxford: Oxford University Press. Pp. 65–75.

Schaefer W (Ed) (1983) *Finalization in Science*. Dordrecht: Reidel.

Schwandt T (2002) *Evaluation Practice Reconsidered.* New York: Peter Lang Publishers.

Scott WG (1995) *Institutions and Organizations*. Thousand Oaks: Sage.

Shadish W and Fuller S (1994) *The Social Psychology of Science*. New York and London: The Guilford Press.

Spaapen J, Dijstelbloem H and Wamelink F (2007) *Evaluating research in context: A method for comprehensive assessment*, 2nd edn. The Hague: Consultative Committee of Sector Councils for Research and Development (COS).

Stark D (2009) *The Sense of Dissonance*: *Accounts of Worth in Economic Life*. Princeton, NJ: Princeton University Press.

Stefik M and Stefik B (2004) *Breakthrough!: stories and strategies of radical innovation.* Cambridge, MA: MIT Press.

Stirling A (2010) Keep it complex. *Nature* 468: 1029–1031.

Stokols D, Fuqua J, Gress J, Harvey R, Phillips H, Baezconde-Garbanati L, Unger J, Palmer P, Clark MA, Colby SM, Morgan G, Trochim W (2003) Evaluating transdisciplinary science. *Nicotine & Tobacco Research* 5, Supplement 1: S21–S39.

Stokols D, Hall KL, Moser RP et al. (2010) Cross-disciplinary team science initiatives: Research, training, and translation. In: Frodeman R, Klein JT and Mitcham C (Eds), *The Oxford Handbook of Interdisciplinarity*. Oxford: Oxford University Press. Pp. 471–493.

Stokols D, Hall KL, Taylor B and Moser RP (2008) The science of team science: Overview of the field and introduction to the supplement. *American Journal of Preventive Medicine* 35(2): S77–S89.

Strathern M (2000a) *Audit Cultures: Anthropological Studies in Accountability, Ethics and the Academy*. London: Routledge.

Strathern M (200b) The tyranny of transparency. *British Educational Research Journal* 26(3): 309–321.

Strathern M (2004) *Commons and Borderlands: Working Papers on Interdisciplinarity, Accountability and the Flow of Knowledge*. Wantage: Sean Kingston Publishing.

Thorngate W, Dawes RM and Foody M (2009) *Judging Merit.* New York, NY: Psychology Press.

Travis GDL and Collins HM (1991) New light on old boys: Cognitive and institutional particularism in the peer review system. *Science, Technology and Human Values* 16(3): 322–341.

Turner S (2000) What are disciplines? And how is interdisciplinarity different? In: Weingart P and Stehr N (Eds), *Practicing Interdisciplinarity*. Toronto: University of Toronto Press. Pp. 25–41.

Weinberg AM (1962) Criteria for scientific choice. *Minerva* 1(2): 158–171.

Weingart P (2000) Interdisciplinarity: The Paradoxical Discourse. In: Weingart P and Stehr N (Eds), *Practicing Interdisciplinarity*. Toronto, University of Toronto Press. Pp. 25–41.

Weingart P (2010) A short history of knowledge formations. In: Frodeman R, Klein JT, Mitcham C (Eds), *The Oxford Handbook of Interdisciplinarity*. Oxford: Oxford University Press. Pp. 3–14.

Whitley R (1984) *The Intellectual and Social Organization of the Sciences.* Oxford: Clarendon Press.

Whitley R (2007) *Business Systems and Organizational Capabilities: The Institutional Structuring of Competitive Competences.* Oxford: Oxford University Press.

Whitley R (2011) Changing governance and authority relations in the public sciences. *Minerva* 49: 359–385.

Whitley R, Gläser J and Engwall L (Eds) (2010) *Reconfiguring knowledge production: Changing authority relationships in the sciences and their consequences for intellectual innovation.* Oxford: Oxford University Press.

Yang K (2011) Emergent accountability and structuration theory: Implications. In: Dubnick MJ and Frederickson HG (Eds), *Accountable Governance: Problems and Promises.* Armonk, NY: M.E. Sharpe. Pp. 269–281.

Yin RK (2003) *Case Study Research. Design and Methods.* Applied Social Research Methods Series, Volume 5. Thousand Oaks, CA: Sage.

Zagzebski, LT (1996) *Virtues of the Mind: An Inquiry into the Nature of Virtue and the Ethical Foundations of Knowledge.* Cambridge: Cambridge University Press.

Ziman J (1991) *Reliable Knowledge: An Exploration of the Grounds for Belief in Science.* Cambridge: Cambridge University Press.

Ziman J (1994) *Prometheus Bound: Science in a Dynamic Steady State.* Cambridge: Cambridge University Press.

Ziman J (1996) Is science losing its objectivity? *Nature* 382: 751–754.

Ziman J (1997) Disciplinarity and interdisciplinarity in research. In: Cunningham, R. (Ed), *Interdisciplinarity and the Organization of Knowledge in Europe.* A Conference Organised by the Academia Europaea. Cambridge, September 24–26, 1997. Luxembourg: Office for Official Publication of the European Communities. Pp. 71–82.

Ziman J (2000) *Real Science: What it Is, and What it Means?* Cambridge: Cambridge University Press.

Appendix 1: The Academy's guidelines for drafting a research plan

(Translated from the Finnish guidelines as defined in August 31, 2007)

The structure of a research plan
A research plan must not exceed 12 pages (Times New Roman 12 pt or corresponding). The plan shall include the following information:

1. Principal investigator (PI), title of research project, site of research

2. Background
   - background and significance of the research, nationally and internationally, including previous research pertaining to the topic
   - how the research project links to other research conducted by the PI or the research team

3. Objectives
   - research objectives
   - hypotheses
   - justifications for how the proposed research ties in with the call and its objectives, if the call has a specific objective (e.g. research program calls)

4. Implementation
   - research methods and research material
   - ethical issues, research permits
   - timetable for the research
   - justifications for the cost estimate specified on the application

5. Researchers and research environment
   - merits and tasks of research team members
   - salary plan for the PI during the funding period, if he/she does not have a permanent contract
   - site of research and any tangible support it offers the project, including available equipment
   - key national and international collaboration and distribution of work: partners, form of cooperation, description of how the project will benefit from the cooperation
   - concrete description of possible work periods abroad, e.g. how the visits or work periods elsewhere contribute to research plan implementation, what are the goals of the visits, are the visits already agreed upon

6. Researcher training and research careers
   - researcher training, including arrangements for teaching and supervision
   - doctoral studies within the PI's team

- promotion of careers of researchers and their planned mobility
- promotion of gender equality within the project

7. Conceivable results
    - expected scientific and societal impact of the research
    - potential for scientific breakthroughs and increasing the capacity for renewal of the science and research
    - applicability and feasibility of the research results
    - publishing of results and raising awareness among potential end-users, the scientific community and the general public

Appendix 2: The evaluation form used by the Academy

(Excerpt from the instructions for evaluating research proposals, given to the expert panelists of 2007)

1. Research plan
   1.1 Scientific quality and innovativeness of the research plan (rating 1–5)
   1.2 Feasibility of the research plan (rating 1–5)
   1.3 The ethical issues involved, if applicable

2. Research environment
   2.1 Competence and expertise of the applicant/research team (rating 1–5)
   2.2 National and international networks of the applicant/research team (rating 1–5)
   2.3 Doctoral and postdoctoral training and development of the research environment (rating 1–5)
   2.4 In the case of a research consortium, significance of the consortium for the attainment of the research objectives

The Academy of Finland requests you to draft written comments for each sub-item. These comments are particularly valuable to the Academy in decision-making. After the funding decisions, the evaluation is also sent to the applicant. The evaluations will give the applicants important guidance for the drafting of their future proposals.

Numerical evaluation of the sub-items is made, as presented above, with ratings ranging from 1 (poor) to 5 (outstanding).

Finally, you are requested to give your overall assessment of the proposal, including the strengths and weaknesses and possible additional comments. Please note that the overall rating should not be a mathematical average of the sub-ratings.

Please rate the application using the scale below. You are encouraged to use the whole scale:

5 = outstanding proposal
4 = excellent proposal, which however contains minor elements that could be improved
3 = good proposal, which contains elements that can be improved
2 = satisfactory proposal, in need of substantial modification or improvement
1 = poor proposal, with severe weaknesses that are intrinsic to the proposed project.

Appendix 3: Interview schedule for funding officers (round I)

1. Background information about you as a funding officer and about the role of funding officers in the evaluation of research proposals

2. The formation of expert panels and the selection of panel members in the evaluation of research proposals

3. Differences and similarities between (uni)disciplinary and multidisciplinary expert panels

4. The role of expert panels' reviews in funding decisions for the proposals, and other funding criteria

5. The organization of the entire evaluation procedure within the research council

6. Interaction and division of labor between research councils (and their staff) in organizing the research evaluation

7. Your personal experiences and views about the evaluation procedure in 2004, reflected from the viewpoint of interdisciplinary research

Appendix 4: Interview schedule for funding officers (round II)

1. The background, rationale, and criteria for organizing the given panel
    - The scope of the panel, including both panelists and proposals
    - The selection of individual panel members
    - The allocation of proposals between panel members

2. The official procedure of the panel meeting
    - Stages, timing, instructions
    - The order and documentation of discussions
    - The formal roles of panel members
    - Formal and informal rules for settling disagreements
    - The criteria and scale of quality

3. Features of the panel discussion and the way in which collective judgments were made
    - Discussions on criteria and on their application in particular cases
    - The length and range of discussions about individual proposals
    - To what extent were proposals or their evaluations compared to each other
    - What is your feeling about the level of unanimity with which collective judgments were made

4. Your impressions of the panel meeting as a whole
    - The atmosphere of the panel meeting; e.g. did the panelists enjoy the meeting
    - What else did the panelists do or talk about, other than discuss proposals, e.g. during the breaks

Appendix 5: Interview schedule for peer review panelists

1. Your areas of expertise
    - How would you describe your area(s) of expertise? How do you perceive yourself as a researcher in your field, in relation to other researchers in the field?
    - To what extent do you perceive your research work as interdisciplinary; how important is this for you, and how do you view the interdisciplinary work of others?

2. Panel's collective expertise
    - How would you describe the collective expertise of your panel; what kind of combination of experts did you form?
    - How was your collective expertise distributed between different panel members; how did the panel members' areas of expertise overlap?
    - Could you describe the evaluation style of each member of the panel? Please compare what properties you appreciated in research proposals to those the other panel members looked for.

3. Roles of panel members
    - Can you describe the roles that different panel members had in panel discussions? Please compare also the role of the two designated evaluators with the role of others.
    - Did the panelists only comment on research proposals within their own fields, or did they comment on other proposals too? How strongly did they express their opinions on different proposals?
    - What is your normative position on the question: Who should participate in evaluating the quality of academic research? Please compare the role of "pure" experts with the role of scholars from neighboring fields.

4. Panel discussion
    - What kind of event would you say the panel meeting was? Describe the atmosphere of the meeting.
    - How did the panel's opinions compare with your individual evaluations of the proposals? Could you describe the process through which the joint opinions were formed in the meeting? How important was the panel discussion for the review process?
    - Could you explain what happened when there was a disagreement between the two preliminary evaluations? Did you settle the disagreement by mutual learning, or by compromising?
    - Did you recognize any direct or indirect disciplinary contradictions during the panel's work?

5. Research proposals
    - What was your impression of the whole set of proposals? Please describe the bulk of proposals in terms of their disciplinary and interdisciplinary approaches.

- Were there some proposals that your panel perceived as notably multi- or interdisciplinary? Can you pinpoint one or two such proposals, and recall how the panel treated them?
- Were there some proposals that the panel was not able to assess very well; why? Could you pinpoint them, and explain what was so distinctive about them?
- If you recognized any disciplinary debates during the panel work, can you specify how the contested questions arose from the respective proposals?

6. Experience of panel work
   - Why did you agree to the request of the Academy of Finland to serve as a member on this expert panel? What do you like most about this kind of evaluation work?
   - Can you tell me about your previous experience of serving on this kind of expert panel? How does this panel of the Academy of Finland compare with your other experiences?

Appendix 6: Operational rules for distinguishing between various types of interdisciplinarity in research proposals

While classifying research proposals in terms of their "type of interdisciplinarity", I developed the following operational criteria that specify the differences between the categories. I applied these criteria whenever a proposal seemed to be on a borderline between two categories. In this table, different types of interdisciplinary research are cross-tabulated in order to illustrate the boundaries between them. Demarcation criteria between each two categories are presented with two opposite descriptions. The first description of each cell refers to the row category, whereas the second description refers to the column category. An empty cell means that the two categories were unlikely to be confused with each other. (The table is on the next page.)

	Encyclopedic multidisciplinarity	**Contextualizing multidisciplinarity**	**Composite multidisciplinarity**	**Empirical interdisciplinarity**	**Methodological interdisciplinarity**	**Theoretical interdisciplinarity**
Disciplinary research	Sub-projects have different scholarly approaches — They lie within the same field	Background and objectives incorporate interests of another field(s) of research — They may incorporate interests of extra-academic stakeholders	Methods, theories and/or data are combined in a new fashion — They constitute a frequently used combination	Data is collected according to several discrete traditions — The triangulation of data is carried out within the tradition of a single research field	Methods are conceptually and methodologically distant from each other or they are applied in a new fashion — They constitute a frequently used combination	The development of concepts/theories/methods is built on more than one antecedent discipline — It relies on only one research tradition established to such a degree that no explicit boundary crossing is needed
Encyclopedic multidisciplinarity		A shared framework — Separate but juxtaposed subprojects	—	—	—	—
Contextualizing multidisciplinarity			Results are integrated — They remain separate	Different sub-projects respond to a shared research question — They only have a shared theoretical framework	—	A novel conceptual framework is developed — An established methodology is applied in a new context
Composite multidisciplinarity				Different sub-projects produce a combination of evidence to solve a shared research problem — Results are aggregated only after different sets of data are interpreted independently from each other	Methods are modified into the new context and the methodology is considered as an integrated whole — Methods are used in an ordinary way	The aim is to produce conceptually novel knowledge — Multidisciplinary approach is selected mainly to supplement the results
Empirical interdisciplinarity					The combination or development of methods includes a bigger interdisciplinary challenge than the combination and interpretation of heterogeneous data — Vice versa	Heterogeneous data are gathered to develop integrative concepts, models or theories — They are gathered to explore their interaction in a particular problem case
Methodological interdisciplinarity						Methodological integration serves broader conceptual development — Methods are integrated in order to study a complex problem