universität wien

DISSERTATION

Titel der Dissertation

Knowing Together: A Social Epistemology for Socio-Technical Epistemic Systems

Verfasserin

Diplom-Psychologin Judith Simon

angestrebter akademischer Grad

Doktorin der Philosophie (Dr. phil.)

Wien, April 2010

Studienkennzahl lt. Studienblatt:	A 092 296
Dissertationsgebiet lt. Studienblatt:	Philosophie
Betreuer:	Ao.Univ. Prof. Dr. DI. Markus Peschl

Table of Contents

List of Figures

List of Tables

Acknowledgements

During the time of writing this thesis I have received a lot of support, personal and institutional support, emotional and intellectual support. Now it is time to express my gratitude to all those who were involved.

Starting with my institutional environment makes me recognize how fortunate I have been during the years of writing this thesis. I was employed at three institutions: the Institute of Philosophy of Science and later the Institute of Philosophy at the University of Vienna as well as the Institute Jean Nicod in Paris. In each of these institutions I have experienced a wonderful combination of intellectual freedom and support that I think is crucial but not always given in academia. First and foremost, I would therefore like to thank my supervisors in these institutional settings in order of appearance: Markus Peschl, Elisabeth Nemeth, Gloria Origgi and Roberto Casati. Each of them has not only contributed greatly to the intellectual environment in these institutions, they also offered their help and their knowledge at various stages throughout this thesis. Special thanks I would like to give to Markus Peschl for supervising my thesis and to Elisabeth Nemeth for offering to act as a second examiner.

Before I continue to other helping hands, I would like to mention another form of institutional support, namely different forms of funding I have received in the course of writing this thesis. Funded by a CEEPUS Teacher Mobility Grant by the Austrian Exchange Service (ÖAW) I was able to spend a month at the University of Ljubljana from May to June 2007 and during this time, the topic of the thesis started to become clearer. In 2008, I was glad to further receive a scholarship for the Stanford University Advanced Graduate Student Exchange Program. This scholarship enabled me to work for six months at Stanford University, where I profited enormously from the University's fabulous research infrastructure as well as the hospitality of both the Forum on Contemporary Europe and the Program in Science, Technology and Society. The University of Vienna also financed the final stages of this thesis with the Research Scholarship F-605, granted for 8 months in 2009. In Paris, I was first the employed by the ANR2008 grant (Agence Nationale de la Recherche, France) CSOSGCAHORS for a Project on "Information Evaluation, Analysis, Organization and Ontologies for Intelligence and Security" and later on by the project LiquidPub, funded by the Future and Emerging Technologies (FET) Programme within the 7th Framework Programme for Research of the European Commission (FET-Open grant number: 213360). I would especially like to thank my colleagues from the LiquidPub project for their support and their patience, especially in the very last months of writing this thesis.

Besides this manifold institutional and financial support, numerous people have offered their intellectual and emotional support in the years of writing. First all of, I would like to thank my colleagues at the University of Vienna, Stanford University as well as the Institut Jean Nicod. In particular I would like to thank Corinna Bath, whose work and the discussions we had have greatly influenced this thesis. Special thanks also go to many fellow PhD-students, whom I have met at in various settings: in different doctoral seminars at the University of Vienna, in the doctoral network *Internet Research,* and in the 2008 cohort of the *Values in Design* program. Many of these fellow PhD-students turned facebook.com, mailing lists and Google Wave into social software with a decided epistemic focus.

Very special thanks I would like to give to those who have read and commented on parts of manuscript. In alphabetical order these were Alban Bouvier, Astrid Mager, Elisabeth Nemeth, Elisabeth Zimmermann, Gloria Origgi, Ingrid Erickson, Markus Peschl, Martin Kusch, Nicole Neuberger, Noah Holtwiesche and Petra Milota. They helped to make the best of this manuscript, yet for its flaws I am the one who is accountable.

Last but not least I would like to thank my family as well as Noah Holtwiesche for their love, support and enduring confidence: to them I dedicate this thesis.

PART 1

1 Introduction

In recent years new applications emerged on the Web which received the labels *Web2.0, social software* or *social media.* The terms were often used interchangeably for a wide variety of applications. Since a clear definition was lacking, examples of such applications would often be listed in lieu of such a definition. Such examples ranged from social networking sites like Facebook.com and MySpace.com to social bookmarking services, like Delicious.com, from online encyclopedias like Wikipedia to recommender systems embedded in commercial websites, from web fora to online photo management and sharing application, like Flickr.com, from prediction markets, like Iowa Electronic Markets, to open innovation systems like Innocentive.com.

The starting point of this thesis was my observation that in many of these applications people are engaged in *epistemic activities*, such as the dissemination, organization or creation of knowledge. They collectively produce *epistemic products*: encyclopedias, classification systems, articles, link collections, calculations, databases, etc. More and more social software applications with such a decided epistemic focus emerged in the last years. They become ubiquitous and given this prevalence, they seem to pervade the epistemic practices of many people – in their everyday quests for information as well as in academia. My interest was sparked to analyze the *epistemological* relevance of these social software applications that have a decided epistemic focus; applications for which I conceived the label *epistemic social software*. From the above mentioned applications, Wikipedia, Delicious.com, recommender systems or Innocentive.com are all clear cases of epistemic social software. Despite their differences, they have one commonality: they are socio-technical systems in which multiple agents are involved in epistemic practices.

To apprehend such systems from an epistemological perspective the first task consisted in finding an appropriate epistemological framework for such an analysis. One crucial characteristic of such systems, and the reason why the term *social* looms so large in their labels, is that *multiple people* are engaged in these epistemic processes together. For this reason, for the fact that the interaction and collaboration between multiple people seem to be the key to understand the epistemic processes within such systems, I chose *social epistemology* as the framework to analyze epistemic social software.

Social epistemology is the philosophical discipline exploring the ways and the extent to which knowledge and epistemic practices are social. As a term, *social epistemology* often refers to a quite specific and narrow field of discourse which evolved since the 1980ies and is rooted primarily in Anglo-American epistemology and philosophy of science. Since the social nature of knowledge has been addressed within philosophy and beyond in numerous ways, broader understandings of social epistemology exist, and there is indeed a considerable overlap to other theoretical developments within and beyond philosophy. The two most notable examples are feminist epistemology as well as different historical and sociological accounts of knowledge and science. Social epistemology, as the mainly analytical, Anglo-American philosophical disciplines that has emerged in the late 1980ies, is at the core of this thesis. However, as shall become obvious in the course of this thesis, understanding the central arguments within social epistemology even in its most narrow conception is impossible without taking into account the broader field of discourse in which these arguments are embedded. Hence, it is rather the *field of discourse around this core that forms the theoretical framework of this thesis.*

I have scrutinized different social epistemologies to see which one could serve best to analyze the socio-epistemic processes taking place within epistemic social software. In the course of these analyses, however, it soon became obvious that none of the existing comprehensive social epistemologies delivers a sufficient framework to analyze epistemic social software. One major reason for this is that many social epistemologies do not account for information and communication technologies (ICT) or technology altogether. And those social epistemologies which address ICT use a model of technology which seems inadequate to me. In particular, ICT is often perceived as a means to only *distribute* knowledge. Since many epistemic social software applications enable the collective *creation* of knowledge, such a narrow focus on distribution is inappropriate. Moreover, the sharp boundary that is drawn between human agents and technologies implies an understanding of technologies as mere tools in the hands of fully rational agents, which also seems inappropriate to understand socio-technical epistemic processes such as the ones characteristic of epistemic social software.

Indicating these shortcomings is less a critique of these social epistemologies per se – after all, they may not have been developed with social software applications in mind. However, it makes such theories insufficient for a more in-depth understanding of the epistemological relevance of epistemic social software.

For these reasons, I started to develop a new socio-epistemological framework for this purpose. This framework crucially builds upon insights obtained from social epistemology. In particular, I have analyzed five major social epistemologies and assessed their merits for an analysis of epistemic social software. Most of them are relevant for certain aspects of socio-technical systems, but none of them offers a comprehensive framework to address the various epistemological issues that emerge in different types of epistemic social software. This is where the framework I propose goes beyond socio-epistemological literature.

Due to the shortcomings of social epistemology indicated above, I have amended my own socio-epistemological framework by insights from other fields. To understand not only the role of technology, but also the relationship between the social, the technical and the epistemic, I rely crucially on literature from the field of *Science and Technology Studies* (STS). Based on my reception of different approaches within STS, I define epistemic social software as *socio-technical epistemic systems consisting of multiple human and non-human agents who interact in different ways for various epistemic purposes*.

My framework is based on a *tripartite classification of socio-technical epistemic system with respect to the mechanisms they employ to close socio-epistemic processes.* Since every epistemic practice is a process, it consists of two temporal occurrences: its initiation and its closure. In this thesis, I argue that analyzing the moment of *closure* is sufficient for differentiating three important types of *epistemic sociality*.

The term e*pistemic sociality* refers to the fact that in all cases there is *more than just one epistemic agent involved in the epistemic process*. Different concepts have been introduced to label and describe the involvement of a multitude of epistemic agents for epistemic tasks: *collective intelligence, wisdom of the crowds, team work, cooperation, collaboration*, to name just a few examples. I use the term *epistemic sociality* as a generic term that encompasses all these notions. The reason for using a generic term here is that each of the before mentioned terms already implies a *specific* way in which the multitude of epistemic agents is organized. Yet those specificities are neither always clear, nor consistent, nor used in the same way by different authors.

Grounded on readings in social epistemology as well as my analyses of epistemic social software, I suggest a classification of three types of epistemic sociality based on the different closure mechanisms employed in different socio-technical epistemic systems**:**

Epistemic Sociality: Integration (ES^I)	Most practices of this type of epistemic sociality are associated with the concept of *collaboration*. To initiate this type of epistemic sociality, epistemic tasks need to be modularized and distributed over different epistemic agents. Once these discrete tasks are accomplished, they need to be *integrated* to achieve the overall epistemic goal.
Epistemic Sociality: Aggregation (ES^A)	Many practices of this second type of epistemic sociality are associated with the concept of the *wisdom of the crowd*. Here, epistemic labor is again distributed over multiple agents, but the mechanism of closure is *aggregation*. I argue that this type of epistemic sociality has its roots in statistical reasoning and is getting reinforced on the Web. Although neither the statistical techniques nor this type of reasoning is new, it is currently on the rise due to its incorporation into information and communication technologies.
Epistemic Sociality: Selection (ES^S)	Instead of *integrating* or *aggregating* epistemic results, in ES^S the main epistemic occurrence of closure is that of *selection.* Analyzing the distribution of cognitive labor over different, competing approaches, over different methods of reaching the *same* epistemic goal, is characteristic for this type of epistemic sociality.

Table 1: Three Types of Epistemic Sociality.

With this classification, I do not aim at reducing the differences between various types of epistemic sociality to their mechanisms of closure. However, I argue that the classification based on this indicator is heuristically fruitful. While each type of epistemic sociality has its own epistemic merits, they all depend on different social, technical and epistemic prerequisites. For instance, while all forms of epistemic sociality depend on *diversity*, the relevance of *independence* differs between the three types. They have different strengths and weaknesses and are optimal for different epistemic tasks. These differences are outlined in detail in the third part of this thesis.

1.1 Goals of this Thesis

The initial goal of this thesis was to provide a comprehensive framework to analyze epistemic social software from a socio-epistemological perspective. Yet, the model I have developed goes beyond this initial goal. It is a new socio-epistemological framework to analyze socio-technical epistemic systems and practices in general. In particular, it is a new socio-epistemological approach that puts something into the focus that has so far been neglected in social epistemology: the technical and its relationship to the social and the epistemic. Since most epistemic practices, in science as much as in everyday-life, are nowadays pervaded by technologies, such a consideration of the role of technologies in these practices indeed seems to be indispensable for *any* social epistemology that aims at being not only normatively appropriate, but also empirically adequate.

The utility of the socio-epistemological framework I propose is twofold. First, providing a schema to distinguish different modes of epistemic sociality is of theoretical interest for social epistemology by offering a more fine-grained and nuanced framework for analyzing socio-epistemic practices on the Web and beyond. Since most socio-epistemological theories have been developed in the context of science, I frequently compare epistemic practices on the Web with those in science. That this comparison does not imply an equation of science and the Web2.0 goes without saying. Nonetheless, I argue that, when it comes to the basic socio-epistemic mechanisms they employ and to the central epistemological – and ethical – questions that have to be raised, they share many commonalities. These commonalities as well as the differences are a recurrent theme throughout this thesis. While the socio-epistemological framework I propose is explicitly developed for the analyses of epistemic social software, it is generic enough to be applied to socio-technical epistemic systems more broadly conceived; systems for which science and epistemic social software are just two very prominent examples.

Differentiating distinct forms of epistemic sociality, outlining their prerequisites as well as their strengths and weakness is crucial for *understanding* the ways in which knowledge is or can be created socially. However, and this is the second goal of this thesis, such a classification can also serve as the theoretical basis for *improving* socio-epistemic practices and systems. Knowledge about the prerequisites, strengths and weaknesses of different social mechanism enables the design and development of socio-technical systems that serve epistemic purposes in a rational, efficient and effective way.

1.2 Methods of this Thesis

The methods employed in this thesis were primarily literature review as well as an analysis of the functionalities of various types of epistemic social software. As a starting point, I have analyzed and described the functionalities of different types of epistemic social software primarily based on their self-description on the Web. Further, I have used additional literature from computer and information science on such systems as well as observations on their usage.

Since social epistemology provides the theoretical core of this thesis, reviewing the major approaches in this field has been of prime importance. These major social epistemologies were Steve Fuller's *Social Epistemology*, Alvin Goldman's *Veristic Social Epistemology*, Miriam Solomon's *Social Empiricism*, Martin Kusch's *Communitarian Epistemology* as well as Helen Longino's *Critical Contextual Empiricism*.

My analysis of these approaches has been twofold. First, I have critically assessed these social epistemologies as epistemological theories addressing the social nature of knowledge in general. Second, I have assessed their merits and shortcomings for providing a theoretical framework to analyze epistemic social software in particular. In doing this, I have developed a grid of important issues to make different social epistemologies comparable and expose their respective strengths and weaknesses. These topics relate to their concepts of knowledge and sociality, the types of epistemic processes that are in focus, as well as the consideration of technology. Beyond these major social epistemologies, I have examined further literature on central socio-epistemological topics which are relevant to analyze epistemic social software. These topics were testimony and trust, consensus formation and distribution of cognitive labor.

Grounded on these readings I was able to provide a map of central debates within the field of social epistemology. I have positioned myself in this field and used this map to develop my own socio-epistemological framework.

Given the interdisciplinary nature of this thesis, I have analyzed literature in various fields beyond social epistemology. The most important insights came from feminist epistemology and philosophy of science as well as from the field or Science and Technology Studies (STS). I have focused on approaches addressing information and communication technologies (ICT), and literature from the field *Values in Design*, a term that refers to the works on the analysis and design of ethically responsible information and communication technologies, has been particularly important.

All these insights obtained from literature review found their way into the development of my socio-epistemological framework. In outlining this model and the three types of epistemic sociality that it distinguishes, I re-assess the examples of epistemic social software outlined at the beginning of the thesis in the light of this framework. I use different types of epistemic social software as support for the validity of my classification and the heuristic fertility of the model proposed. I conclude each chapter of the different types of epistemic sociality with recommendations concerning the analysis and design of epistemic social software conceived as socio-technical epistemic systems consisting of multiple entangled human and non-human epistemic agents.

1.3 Structure of this Thesis

The thesis is divided into three major parts. In the first part I apprehend my empirical object of analysis: epistemic social software.

In Chapter 2, I explain why I introduce the term *epistemic social software* and discuss to what extent a distinction between epistemic and non-epistemic social software can be made. Defining epistemic social software is difficult in so far as the term social software itself is ill-defined. Hence, in lieu of a proper definition, I portray different examples of epistemic social software. Having screened the field of epistemic social software, I chose examples which prototypically depict different ways in which the social, the technical and the epistemic are entangled on the Web. These major examples are Wikipedia, Delicious.com, Innocentive.com, and Recommender Systems. While I refer to other examples of epistemic social software throughout the thesis, these four examples serve as points of reference to which I return frequently. The portrayal of these applications in Chapter 2 is primarily based on their self-description on the Web and remains rather cursory, because they are reassessed in more detail in the light of my own socio-epistemological framework in Part 3 of the thesis.

In Chapter 3 I argue that instead of understanding epistemic social software as mere tools or applications, they should be conceived as *socio-technical epistemic systems consisting of multiple human and non-human agents who interact for various epistemic purposes*. To support such a conception of epistemic social software I take a brief detour through the field of Science and Technology Studies (STS). I outline and assess the merits of the major STS-approaches for the analyses of epistemic social software. These approaches include the *Social Shaping of Technology* (SST), the *Social Construction of Technology* (SCOT) Actor-Network Theory (ANT)), the feminist approaches of Donna Haraway, Karen Barad and Lucy Suchman and finally different approaches within the more pragmatically oriented field of *Values in Design*. The key insights of this chapter for the analyses of epistemic social software conceived as socio-technical epistemic systems concern the following issues: the *entanglement of the social, the technical and the epistemic*; the *relationship between human and non-human actors*, a *performative understanding of epistemic practices and systems*, and the *relationship between values and technologies.*

Chapter 4 completes the first part of this thesis with a summary of the main arguments and the conclusions to be drawn about epistemic social software conceived as *socio-technical epistemic systems in which multiple human and non-human agents interact for epistemic purposes*.

In the second part of this thesis, I introduce social epistemology as the major theoretical framework, outline central debates and position myself within this field of research. In Chapter 5 I delineate the field of social epistemology as conceived in this thesis. Although the term itself was coined in the library science in the 1950ies, social epistemology has later become to refer to a mainly Anglo-American, analytically oriented endeavor that emerged in the 1980ies. My field of reference is the *discourse around this narrow conception of social epistemology forms*. I then portray the five major comprehensive social epistemologies proposed by Steve Fuller, Alvin Goldman, Miriam Solomon, Martin Kusch and Helen Longino in some detail.

While in Chapter 5 I remain quite neutral in my portrayal of these different approaches, I scrutinize them in the next chapter. Thus, Chapter 6 is devoted to an assessment of social epistemology as a theoretical framework to analyze epistemic social software. In this chapter I analyze additional socio-epistemological literature on topics which are crucial for the assessment of epistemic social software. These topics of concern are testimony, trust, consensus formation and the distribution of cognitive labor. Based on analyses of the theories portrayed in Chapter 5 as well as this additional literature, I outline central debated in the field of social epistemology and position myself with respect to these debates. In a next step I outline initial attempts to analyze social software from a socio-epistemological perspective. There have been only few socio-epistemological analyses of social software and although they deliver important insights, they can only be considered to be starting points for a more comprehensive assessment of the epistemic merits and shortcomings of different types of epistemic social software.

In Chapter 7 I summarize the main arguments as well as my stance towards the issues raised and draw conclusions concerning the implications of my analyses for the development of a comprehensive socio-epistemological framework for the analysis of epistemic social software.

In Part 3 of the thesis, I describe my socio-epistemological framework to analyze epistemic social software, i.e. social software whose *primary* purpose lies in the creation, dissemination or evaluation of knowledge. I argue that my framework goes beyond previous socio-epistemological analyses of ICT by emphasizing the processes of collective knowledge *creation* taking place in social software understood as socio-technical epistemic systems instead of focusing on distribution of knowledge only. This framework is built upon the insights obtained from the different fields of research addressed throughout this thesis.

In Chapter 8 I outline the basics of my framework. I show how the various insights are combined in my conception of epistemic social software as socio-technical epistemic systems and what implications this conception has for a socio-epistemological analysis of such systems. In particular, I make use of Helen Longino's tripartite notion of knowledge as content, practices and cognitive agency to show how knowledge as content is input and output of knowledge-productive practices conducted by entangled and situated epistemic agents. The production of knowledge depends on the temporary closure of socio-epistemic processes. I distinguish three generic mechanisms of such closure, namely *integration, aggregation* and *selection* and use them to develop a comprehensive socio-epistemological framework based on this triple classification of socio-technical epistemic systems with respect to the mechanisms of closure they employ. Before introducing each type of epistemic sociality, *ES:Integration (ES^I), ES:Aggregation (ES^A)* and *ES:Selection (ES^S)* in detail, I portray two additional sources of inspiration for my model: Yochai Benkler's analyses on commons-based peer production as well as James Surowiecki's thoughts on the wisdom of the crowds.

In the next three chapters I outline these three types of epistemic sociality in detail. In each chapter I describe the key characteristics of each type, offer examples from science and the Web, outline key issues that need to be addressed for any socio-epistemological analyses of systems employing the respective mechanism of closure, draw conclusions concerning the strength, weaknesses and prerequisites for each type and conclude the description with recommendations concerning the analyses and design of such systems.

Chapter 9 is devoted to the first type of epistemic sociality (ES) which I distinguish: ES:Integration (ES^I). After epistemic tasks have been distributed over multiple epistemic agents, these tasks need to be integrated for the collective epistemic result. Using examples from the Web and from science, I outline socio-epistemological considerations of ES^I and emphasize several key issues that need to be addressed when analyzing or designing socio-technical epistemic systems that exploit ES^I. These topics concern the structure of epistemic tasks, as well as diversity, trust, authority and reputation. The chapter ends with some conclusions concerning the socio-epistemological analysis and design of epistemic social software.

In Chapter 10 I outline the second type of epistemic sociality, ES:Aggregation (ES^A) following the same schema. I describe how ES^A differs from other types of epistemic sociality and provide different examples of systems exploiting ES^A. I analyze socio-epistemological perspectives on the epistemic merits of aggregational mechanisms and outline the key topics, before I draw some conclusions and provide recommendations concerning systems employing ES^A.

In Chapter 11 I outline the third and final type of epistemic sociality, ES:Selection (ES^S). Instead of *integrating* or *aggregation* epistemic results provided by different epistemic agents this type of epistemic sociality consists in a selection. ES^S hence differs from the other two types of epistemic sociality, because singular results do not have to be re-combined. Moreover, ES^S is particularly suited to elucidate a crucial aspect of my three types of epistemic sociality that has not been addressed so far: the possibility to combine and nest different types of epistemic sociality in socio-technical epistemic system and processes. After outlining the central characteristics of ES^S and how it differs from ES^I and ES^A, I refer to socio-epistemological analyses on diversity and the distribution of epistemic labor in science and outline different key topics which I elucidate with examples from the Web and from science. As was the case in the other chapters, I conclude my analyses with some conclusions and recommendations concerning the analysis and design of socio-technical epistemic system that employ selection as a mechanism of closing socio-epistemic processes.

In Chapter 12 I conclude this thesis by summarizing the key insights, outlining its limits and open up new possibilities for future research. While I hope to have shown the merits of my interdisciplinary approach throughout this thesis, I highlight some of the problems that it entails. Interdisciplinary research, especially if conducted by a singular epistemic agent, always entails the risk of falling short of interdisciplinary standards, because it needs to suffice several disciplinary standards at once. The disciplines touched in this thesis are philosophy, science and technology studies, as well as computer and information science. Even within each discipline one can get easily lost. Not to get lost in all of them makes it necessary to rigorously omit the vast majority of research within each field. I hope to have made wise choices in what to include and what to omit. Yet this judgment will ultimately reside in the eyes of the community of my readers.

2 Social Software: A Cursory Introduction

In the last years, certain developments in information and communication technologies (ICT) have received a great deal of attention not only within technological communities, but also in economic (e.g. Benkler 2002; Benkler 2006, Tapscott and Williams 2006), legal (e.g.Lessig 2001, Lessig 2006, Koepsell 2003, Gasser 2006, Palfrey and Gasser 2008), or political discourses. (e.g. Sunstein 2002). Those developments I am referring to have been labelled *social software, Web2.0, social media, or social web*. In the following the terms are used interchangeably.

As Luciano Floridi has noted, no clear definition concerning the scope or nature of social media or Web2.0 has been proposed yet (Floridi 2009). Most portrayals of social software rather list applications considered prime examples instead of defining the term. Some of the most well-known examples of social software are Wikipedia, a collaborative online encyclopedia; Delicious and CiteULike, as social bookmarking services; flickr.com and youtube.com as tools to share photos and videos, facebook.com and myspace.com, as general social networking sites, LinkedIn and Academia.edu, as social networking sites for business or academia. Given their focus on networking, sharing or collaboration, the *sociality* of these tools and applications should be self-evident. However, also other tools, which are less obviously "social" fall into the category of Web2.0 or social software. Recommender systems, employed on commercial websites like amazon.com, which suggest interesting products to users also harvest some form of "wisdom of the crowd" by generating recommendations based on information obtained from multiple users. Further tools for peer-to-peer sharing of information or music also fall into the category of social software, such as Napster, Gnutella or filestube.com to use a more recent example.

Social software is omnipresent these days with more and more applications and tools being developed each week. In their ubiquity, social software applications have a profound effect on many societal domains and as a consequence, numerous theoreticians from economy, law, education and politics such as the ones introduced above, have analyzed the relevance of social software in their respective domains. In contrast to this excessive engagement in those fields, the *epistemological* relevance of social software has only quite recently started to receive any attention. This thesis aims at filling this gap by providing a *comprehensive theoretical framework for analyzing social software* *from a socio-epistemological perspective*. Social epistemology is a philosophical discipline dealing with ways and the extent to which knowledge is social and it forms the theoretical frame of reference for my analyses. The field of social epistemology is outlined in the second part of this thesis, whereas my own model is developed in the third part, but before turning to the theoretical framework, I start by introducing the phenomena that have ignited these analyses in the first place in some detail.

At first I introduce some definitions of the terms Web2.0, social media or social software. I argue that no clear definition of this phenomenon has been provided yet. Hence I follow a diametrical approach by given examples instead of a definition to introduce the field. Given the overabundance of tools and the fast pace of developments, I do not aim at giving an exhaustive list. Instead I use several *maps* of social software that I found on the Web: various visualizations of the magnitude and variety of social software have been created by different people and I use them to give an idea about the scope of applications that fall into this category of software. After having opened up the field I narrow it again by restricting my focus to *epistemic social software*, i.e. social software whose primary purpose is epistemic. I depict some examples of such *epistemic social software* in more detail, namely Wikipedia, Delicious, InnoCentive and Recommender Systems. I have chosen these examples, because these systems paradigmatically depict *different ways in which the social, the technical and the epistemic are entangled on the Web*. Since I return to these examples in detail in Part 3 again, their portrayal in this chapter remains rather cursory and descriptive.

2.1 An Ill-Defined Success

In a comparison between Web2.0 and Web3.0, Floridi notes that while Web3.0, also known as *Semantic Web* is a well-defined mistake, Web2.0 – or social software is an ill-defined success. He states that Web 2.0 refers to a "loose gathering of a wide variety of family-resembling technologies, services, and products" (Floridi 2009: 31). Nonetheless, he uses the Wikipedia definition of Web2.0 as a starting point for his analyse. The definition he cites reads as follows:

> "Web 2.0 concepts have led to the development and evolution of web-based communities and its hosted services, such as social-networking sites, video sharing sites, wikis, blogs, and folksonomies. The term became notable after the first O'Reilly Media Web 2.0 conference in 2004. Although the term suggests a new version of the World Wide Web, it does not refer to an

update to any technical specifications, but to changes in the ways software developers and end-users utilize the Web. According to Tim O'Reilly: "Web 2.0 is the business revolution in the computer industry caused by the move to the Internet as platform, and an attempt to understand the rules for success on that new platform."'[1]

Further Floridi cites O'Reilly in defining Web 2.0 as,

> "the network as platform, spanning all connected devices. Web 2.0 applications are those that make the most of the intrinsic advantages of that platform: delivering software as a continually-updated service that gets better the more people use it, consuming and remixing data from multiple sources, including individual users, while providing their own data and services in a form that allows remixing by others, creating network effects through an "architecture of participation", and going beyond the page metaphor of Web 1.0 to deliver rich user experiences" (Floridi 2009: 31).

At the time of writing this chapter, the same entry in Wikipedia to which Floridi referred reads quite similarly, although certain shifts are recognizable:

> "The term "Web 2.0" (2004–present) is commonly associated with web applications that facilitate interactive information sharing, interoperability, user-centered design, and collaboration on the World Wide Web. Examples of Web 2.0 include web-based communities, hosted services, web applications, social-networking sites, video-sharing sites, wikis, blogs, mashups, and folksonomies. A Web 2.0 site allows its users to interact with other users or to change website content, in contrast to non-interactive websites where users are limited to the passive viewing of information that is provided to them.[2]

Reference to interactive and collaborative processes was strengthened and the definition now entails numerous examples, such as wikis, blogs, etc. When looking up related notions of social software, social media or social web on Wikipedia, the following definitions can be found.

> "Social software encompasses a range of software systems that allow users to interact and share data. This computer-mediated communication has become very popular with social sites like MySpace and Facebook, media sites like Flickr and YouTube as well as commercial sites like Amazon.com and eBay."[3]

In the article on *social software*, the following types of social software are listed: instant messaging, text chat, Internet forums, blogs, wikis, collaborative, real-time editors, prediction markets, social network services, social network search engines, deliberative social networks, commercial social networks, social guides, social bookmarking, social

---

[1] http://en.wikipedia.org/wiki/Web_2.0, [date of access: 31.10.2008], quoted from (Floridi 2009: 31).
[2] http://en.wikipedia.org/wiki/Web_2.0, [date of access: 11.01.2010].
[3] http://en.wikipedia.org/wiki/Social_software [date of access: 11.01.2010].

cataloguing, social libraries, social online storage, virtual worlds as well as other specialized social applications, such as Project management and e-learning applications.

Social media in turn has been defined as "a group of Internet-based applications that build on the ideological and technological foundations of Web 2.0, and that allow the creation and exchange of user-generated content" (Kaplan and Haenlein 2010). And the Wikipedia entry on social media further states that it is"media designed to be disseminated through social interaction, created using highly accessible and scalable publishing techniques. Social media uses Internet and web-based technologies to transform broadcast media monologues (one to many) into social media dialogues (many to many). It supports the democratization of knowledge and information, transforming people from content consumers into content producers."[4]

Even without offering more definitions of Web2.0, social media and social software three things should be obvious. First, the terms Web2.0, social media and social software are clearly related, although they are not always conceived congruently. Secondly, the terms refer to similar tools and applications, although the fringes of what is considered Web2.0, social media and social software are fuzzy. Third, not only the fringes, but also the core of Web2.0, social media and social software remains vague. The most that can be said is that it is usually *a multitude of epistemic agents that is interacting, sharing or co-producing content.*

Hence, Floridi seems to be right in stating that Web2.0 is ill-defined. But he also states that it is a success story, an appraisal that many theoreticians from a diversity of backgrounds share. What is the reason for the success of social software? And can this reason give us a hint at the specificities of social media even if no "watertight list of necessary and sufficient conditions that should qualify something as Web 2.0" may be possible (Floridi 2009: 31)?

2.2 A Matter of Family Resemblance

Siding with Floridi that finding necessary and sufficient conditions to define Web2.0 might indeed be impossible, I take a diametrical approach by providing examples of applications that are usually considered to be social media, social software or Web2.0. Given the abundance of tools that are available already and the speed by which new

[4] http://en.wikipedia.org/wiki/Social_Media [date of access: 11.01.2010].

tools are being developed, I cannot deliver an extensive lists of applications. Instead, I depict several maps of social software applications that I have found on the Web by using a search engine. In these visualization, which were posted on personal blogs or on Flickr.com, an online photo management and sharing applications, people aimed to either give an *overview* over different types and examples of social software or to describe their own *idiosyncratic usage* of social software as a personalized map.

Figure 1: Conversation Prism by Brian Solis[5]

The first map above is the Conversation Prism, provided by Brian Solis and JESS3. As indicated by the title, *conversation, the art of listening, learning and sharing* forms the core of this map. They have classified Web2.0 into 22 types of applications. Some of the classes are defined by the type of content (e.g. music, video, pictures), some by the type of technology (e.g. blogs, wikis) others by a combination of types of actions and types

[5] http://www.briansolis.com/2008/08/introducing-conversation-prism/ [date of access: 10.02.2010]

of content (e.g. video aggregation), still others refer to temporal aspects (e.g. LiveCasting). Within each of the classes numerous examples of Social software are listed, such as Facebook or Hi5 as two examples of social networks, last.fm and Pandora as music platforms, etc.

Another visualization I have chosen because it gives an overview over available tools is provided by Fred Cavazza. He also lists numerous examples but uses different classification system. Listing different social platform, he distinguishes four major functions of social media for classification: expressing, networking, sharing and gaming. Each of the four functions is then subdivided: For instance, expressing involves publication, discussion and aggregation; networking involves search, niche, business-to-business, mobile and tools; sharing concerns either content, products or places. For gaming finally also five classes of games are distinguished.

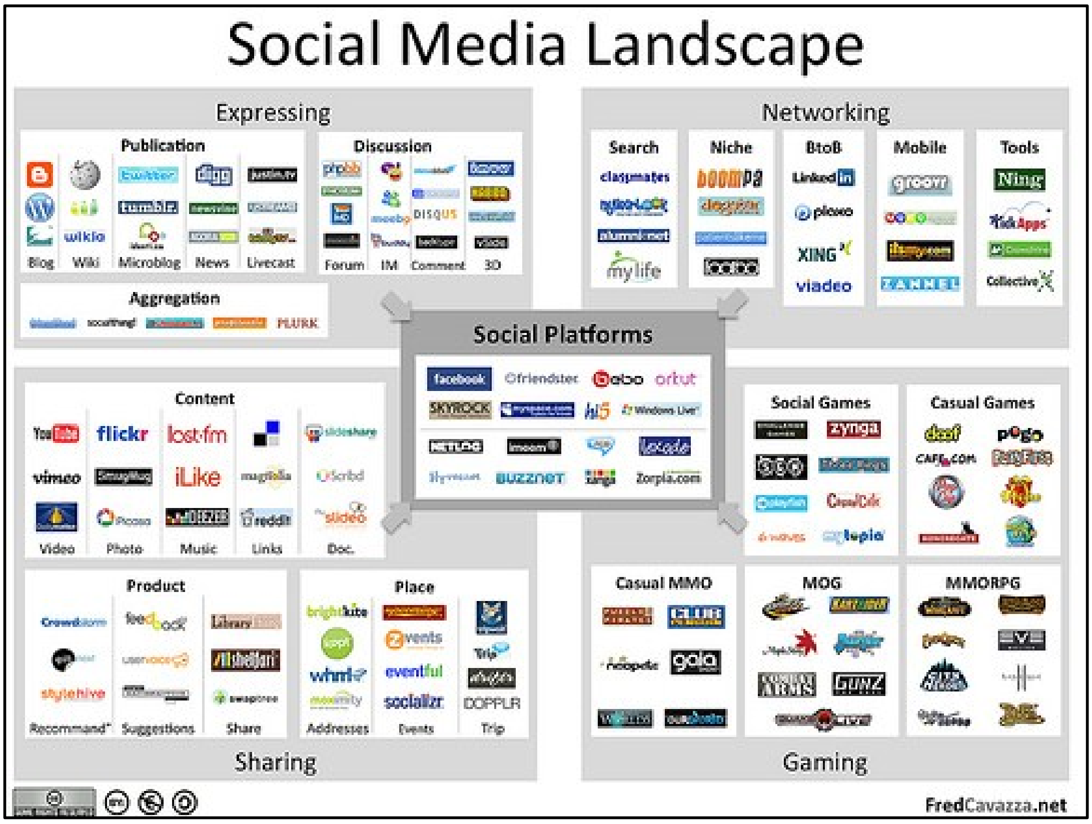

Figure 2: Social Media Landscape by Fred Cavazza[6]

[6] http://www.flickr.com/photos/fredcavazza/3428921418/ [date of access 10.02.2010]

Another visualization provided by Laurel Papworth and Gary Hayes lists the same tools as already shown in the previous two displays, but sort them along the dimensions of *involvement, creation, discussion, promotion* and *measurement*.

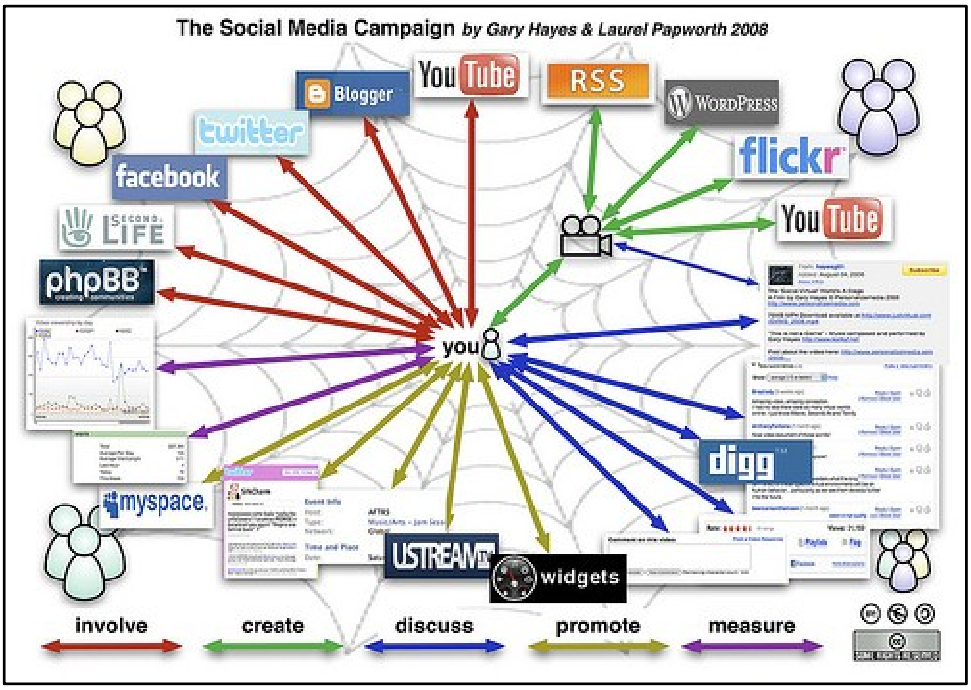

Figure 3: The Social Media Campaign by Laurel Papworth & Gary Hayes[7]

The next examples provide rather idiosyncratic maps of usage, in which the user is in the center and sorts those social media that he uses along different criteria. Frank Da Silva's map is organized around content (photos, videos), functions (search, bookmarking, mircoblogging), and type of application (blog, email).

[7] http://www.flickr.com/photos/garyhayes/2973684461/ [date of access 10.02.2010]

Figure 4: My Cyber Social Map by Frank Da Silva[8]

In the final example of a personalized map in which the user is central social media is sorted along the four actions of *create, communicate, consume* and *connect*.

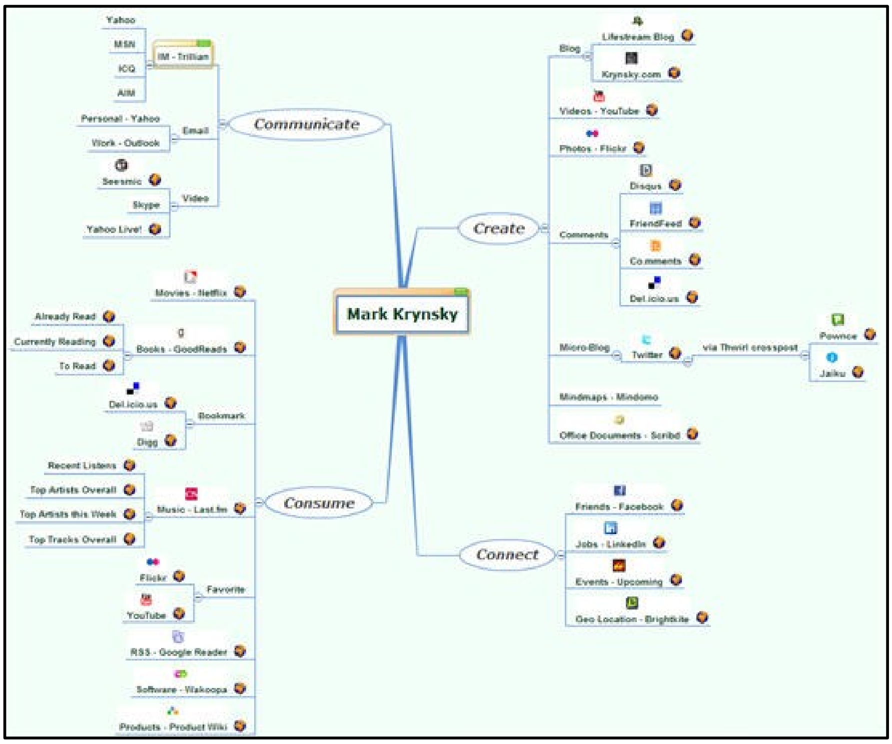

Figure 5: Personalized Map of Social Media by Mark Krynsky[9]

[8] http://socialgraphcentral.wordpress.com/2008/04/03/frank-da-silvas-social-map/ [date of access: 10.02.2010]

There are several reasons why I have portrayed several of these maps. First of all, all maps have in common that they show a plenitude of social software or Web2.0-applications that are available. However, they also differ in numerous respects. While some of them are *idiosyncratic views* describing personalized usage of different tools where the user is in the center, others rather provide a *general overview*. Moreover, the maps show that *different classifications* of such social software are possible depending on the dimensions emphasized. These classifications do not only differ in their *granularity*, they also use - and often confound - different criteria for classification: type of content, functions, types of underlying software, etc. Hence, there are numerous ways in splitting the field of social software or social media. The criterion which I have chosen to distinguish different types of social software concerns the question of whether their primary purpose is *epistemic or not*.

2.3 Epistemic Social Software

After having outlined the plethora of tools and applications that have been labelled social software of Web2.0, in the following section I narrow the focus again. Social software is used for numerous purposes: to connect to people one either already knows or wants to get to know; to communicate; to share information, music, pictures, videos, to create content collectively. Many of the purposes listed above have an *epistemic dimension*. Sharing information, creating an encyclopedia, finding collaborators for a research project are among the purposes that are clearly related to knowledge creation and dissemination. In some applications, such as dating platform or p2p-networks to share music, epistemic purposes may be less relevant.

The focus of this thesis lies on *social software whose primary purpose is epistemic*, i.e. in which knowledge creation, dissemination or evaluation loom large. I have labelled them *epistemic social software.* I have chosen different examples of epistemic social software, which I depict in same detail below. These examples were chosen because they emphasize different socio-epistemic practices, different possibilities in which a multitude of users in interaction with each other and the technological infrastructure engage in the creation, dissemination or evaluation of epistemic content.[10]

[9] http://lifestreamblog.com/tag/social-media/ [date of access: 10.02.2010]
[10] My epistemological foundations, the definition of knowledge adapted, etc. is outlined in the second part of this thesis. For the moment it suffices to know that knowledge is a success term that depends on being

Put differently, I do not focus on all other aspects of social software, such as meeting new people for the sake of meeting new people, listening to music for the sake of listening to music, etc. The specification "for the sake of" implies a caveat. Of course the distinction between social software whose purpose is primarily epistemic as opposed to those whose primary purpose lies in communication or networking or entertainment is somewhat artificial. Just taking a look at my own usage of social software reveals the difficulty of this distinction. Although *facebook.com* may be a primarily non-epistemic application, I frequently use it to ask colleagues and friends for recommendations on literature, methods or technical advice. This is the case because a substantial percentage of my facebook-friends are indeed people I know from my research. Another example is *Skype*, a voice over IP software. I use it to stay in touch with my parents, my partner and friends but also to coordinate research activities and communicate with colleagues. I use *Delicious*, a social bookmarking service, to save links for entertainment and information about Paris, cooking and hiking as well as links that are related to my research. I use it to save information for me personally, but also to share it with my colleagues by using a shared tag for content related to our research.

These examples of my own idiosyncratic practices show that the usage of social software is not entirely prescribed by the software itself. Social software is mostly not designed to be *purely epistemic* or *purely non-epistemic*. Rather, the extent to which they are used for epistemic purposes depends on the users. This aspect of social software can be grasped with the term *interpretative flexibility* Pinch and Bijker 1987: people use social software for different purposes, they can *appropriate* it according to their needs. Hence the distinction between epistemic and non-epistemic social software resides in the *usage* and not primarily in the *design* of the applications. Clearly, design plays a role because it makes it more likely that applications are used for one or another purpose. But in the end, usage depends on the users – and there probably is considerable variance in the usage of different tools between different users.[11]

Despite these caveats, I think the distinction between tools who are primarily epistemic and those who are not is a useful on for the purpose of this thesis, because it enables to

acknowledged to be knowledge by a community of knowers. Hence, knowledge only starts to exist when at least two people are engaged in the process. Nonetheless, individuals can produce epistemic content, which is a general term for intellectual products that may or may not have been vetted as knowledge.

[11] The issue of appropriation, interpretative flexibility and the role of designers are discussed in more detail in the next chapter.

narrow the field and focus on relevant questions. In the following I depict some exemplary cases of epistemic social software in some detail. The examples I have chosen are Wikipedia, Delicious, Innocentive.com, and recommender systems, because they elucidate paradigmatic cases in which the social, the technical and the epistemic are entangled on the Web. In the third part of this thesis I develop a model to classify different social software applications and provide a theoretical framework for their analyses. I then return to these examples again and therefore their portrayal in this chapter remains rather descriptive.

2.4 Prime Examples of Epistemic Social Software

2.4.1 Wikipedia: Creating Content Together

Wikipedia is one of the greatest success stories of the Web - and it is a prime example of epistemic social software. An army of more than 85.000 volunteers[12] has provided a source of information that millions of people trust and use regularly in order to acquire knowledge on all sorts of topics. Taking a look at its self-description reveals its magnitude and main characteristics.

> "Wikipedia [...] is a multilingual, web-based, free-content encyclopedia project based on an openly-editable model. The name "Wikipedia" is a portmanteau of the words wiki (a technology for creating collaborative websites, from the Hawaiian word wiki, meaning "quick") and encyclopedia. [...] Wikipedia is written collaboratively by an international group of volunteers. Anyone with Internet access can write and make changes to Wikipedia articles. Users can contribute anonymously, under a pseudonym, or with their real identity if they choose."[13]

According to their self-description Wikipedia attracts about 65 million visitors a month. At the time of writing this chapter, English-language edition of Wikipedia has 11.639.040 registered users and consisted of 3.190.515 content pages, (19.416.335 pages including talk pages, redirects, etc.) and 867,850 uploaded files.[14] However, the English-language edition of Wikipedia is only the largest edition. Wikipedia exists in 272 languages.[15] Below you can find a screenshot of the start page of Wikipedia and

---

[12] Unless otherwise noted the information on Wikipedia is taken from its self-description available at: http://en.wikipedia.org/wiki/Wikipedia:About [date of access: 12.02.2010]
[13] http://en.wikipedia.org/wiki/Wikipedia:About [date of access: 11.02.2010]
[14] http://en.wikipedia.org/wiki/Special:Statistics [date of access: 11.02.2010]
[15] http://meta.wikimedia.org/wiki/List_of_Wikipedias [date of access: 11.02.2010]

further below some information about the size of the ten largest Wikipedia editions based on data from 2007.

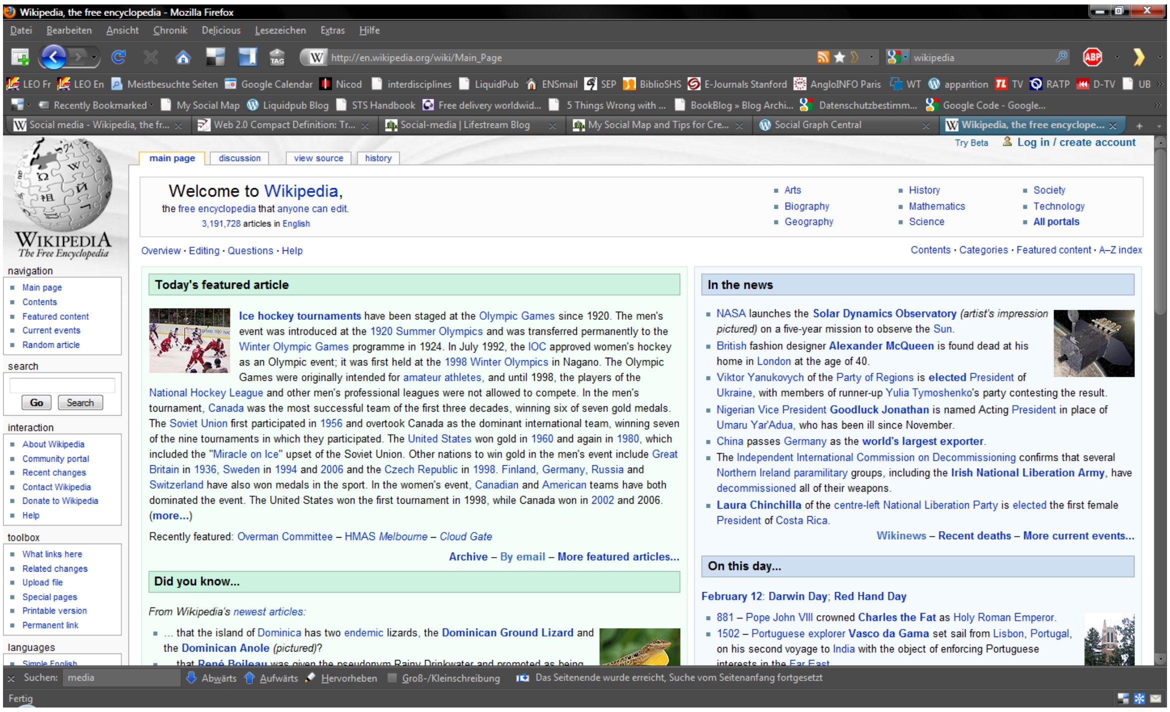

Figure 6: Start Page of the English-Language Wikipedia[16]

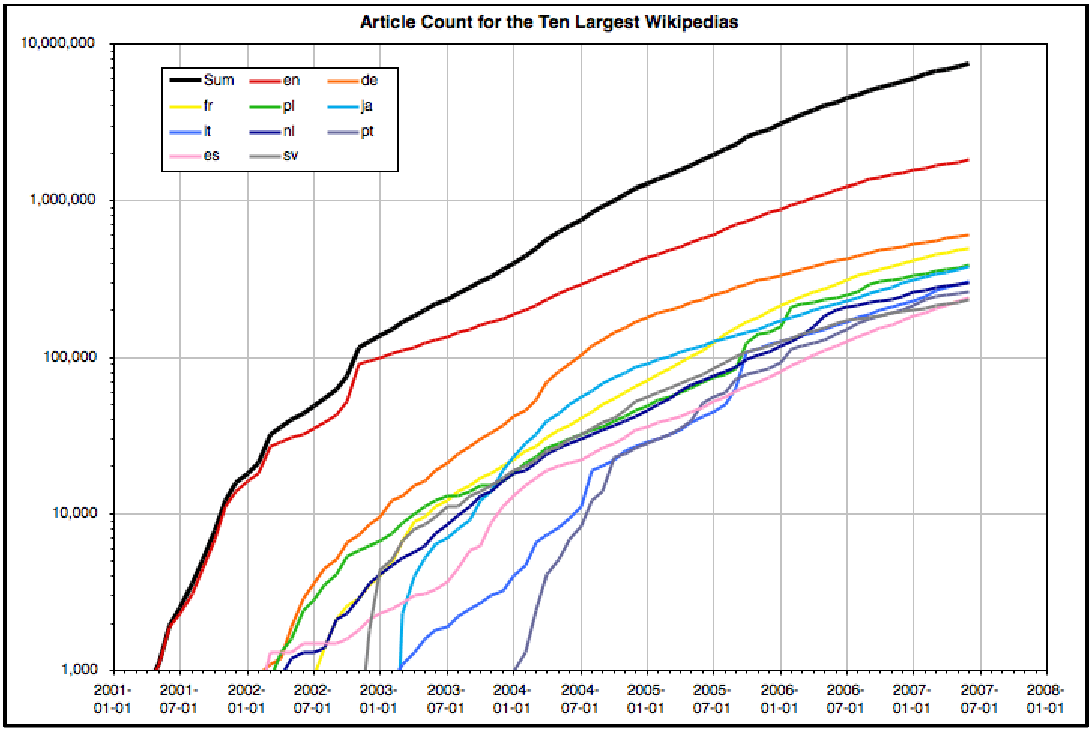

Figure 7: Article Count for the Ten Largest Wikipedias until 2007[17]

[16] http://en.wikipedia.org/wiki/Main_Page [date of access: 11.02.2010]

The *technical basis* of Wikipedia is a wiki. A wiki basically is a website that allows users to collaboratively create and edit web pages using a web browser. The first wiki has been developed by Ward Cunningham in the mid-nineties as a tool to support collaboration between programmers.[18] He defined a wiki as "[t]he simplest online database that could possibly work. Wiki is a piece of server software that allows users to freely create and edit Web page content using any Web browser. Wiki supports hyperlinks and has a simple text syntax for creating new pages and crosslinks between internal pages on the fly."[19]

By now numerous wiki software applications exist and MediaWiki, the basis of Wikipedia, is just one of them. A main characteristic of a wiki is that multiple users can use it simultaneously to add or change content. As opposed to the temporal structure of blogs, wikis do not have such an intrinsic structure, but can rather be formed according to specific purposes. Hence the structure of Wikipedia is just one instantiation of a wiki.

Two important features of Wikipedia are the *history and the discussion sites* which are linked to each website. The *discussion page* is allows to have conversation around the creation of content. By displaying the amount of controversy around an article, it can be used as another indicator to assess the quality of the information provided in the article. Via the *history page,* all revisions can be tracked and attributed to either registered users or the IP-numbers of unregistered users. Moreover, every change can be undone by a simple mouse click. These features of the history page are of particular importance for quality control on Wikipedia. Not only can errors be attributed to users, the ease of revision is also intended as a tool to counter vandalism, i.e. the intentional distortion of information on Wikipedia. Both the history and the discussion pages are an attempt to increase transparency about the production processes in Wikipedia.

---

[17] http://en.wikipedia.org/wiki/Wikipedia:Size_of_Wikipedia [date of access: 11.02.2010]
[18] http://c2.com/cgi/wiki?WikiHistory [date of access: 12.02.2010]
[19] http://www.wiki.org/wiki.cgi?WhatIsWiki [date of access: 12.02.2010]

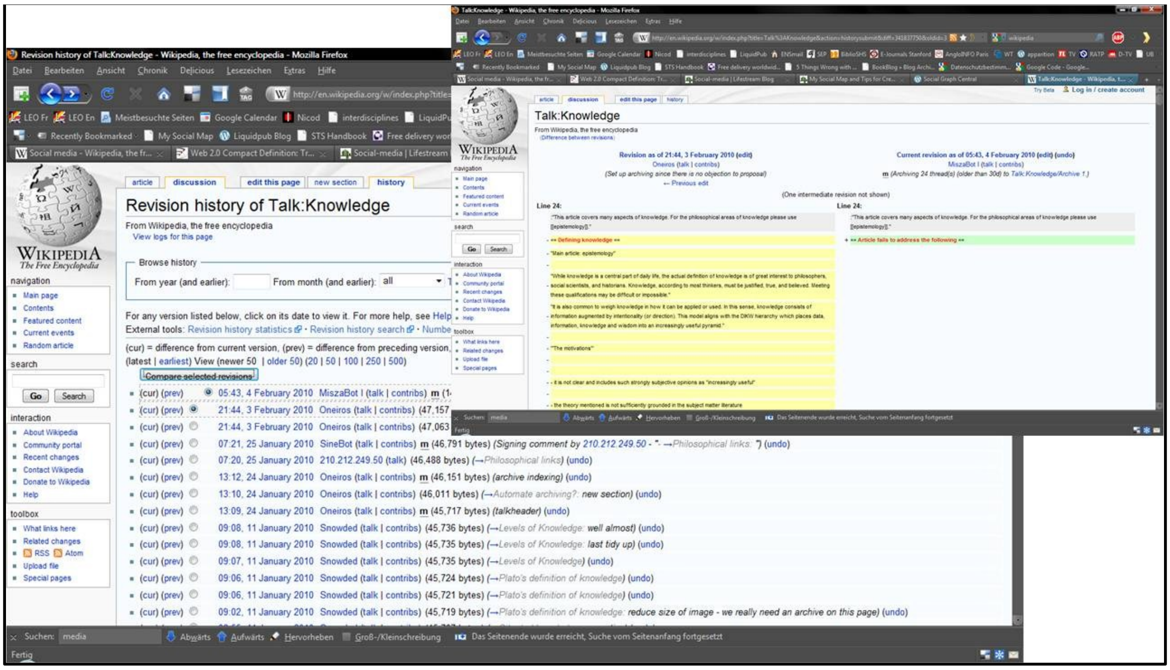

Figure 8: Screenshots of the History Pages of the Entry on "Knowledge"[20]

The overarching *goal* of Wikipedia to provide a free-content encyclopedia that covers existing knowledge of the world. Attached to this notion of an *encyclopedia*, there are restrictions and specifications with respect to content allowed on Wikipedia. In specific, there are three core content policies: *Neutral Point of View, Verifiability* and *No Original Research.*

The neutral point of view policy implies that "[a]ll Wikipedia articles and other encyclopedic content must be written from a neutral point of view, representing fairly, proportionately, and as far as possible without bias, all *significant* views that have been published by reliable sources."[21] The emphasis on reliable sources leads to the second policy, which concerns the verifiability of information. All articles should provide sources for their information and link to this source via citation. From an epistemological perspective it is interesting to note that "[t]he threshold for inclusion in Wikipedia is verifiability, not truth—what counts is whether readers can verify that material added to Wikipedia has already been published by a reliable source (see below), not whether we think it is true."[22]

[20] http://en.wikipedia.org/w/index.php?title=Talk:Knowledge&action=history [date of access: 11.02.2010]
[21] http://en.wikipedia.org/wiki/Wikipedia:Neutral_point_of_view [date of access: 12.02.2010]
[22] http://en.wikipedia.org/wiki/Wikipedia:Verifiability [date of access: 12.02.2010]

Another less straightforward implication concerns the omission of original research. This aspect however, is directly linked to requirement of verification. Since only knowledge which can be verified by other sources should be included in Wikipedia, it is stated that “[a]rticles may not contain any unpublished theories, data, statements, concepts, arguments, or ideas; or any new interpretation, analysis, or synthesis of published data, statements, concepts, arguments, or ideas that,[...] would amount to a "novel narrative or historical interpretation."[23]

Wikipedia can be described as a large-scale collaboration of volunteers. However, while it is often claimed that everybody can participate, this participation is bound to some prerequisites: access to a computer and the Internet, technical skills, etc., and these prerequisites are clearly neither equally distributed within countries nor between them. Moreover, the large majority of users rather absorb information without actively contributing to the encyclopedia. Hence, openness is a matter of degree and while the entrance barrier for Wikipedia certainly is low, there are prerequisites which have to be met; there are skills and technologies that one needs to possess to be able to participate. Despite these caveats, the idea of open participation is a central characteristic of Wikipedia and implies that there is no screening or selection based on competency (or honesty): anyone with Internet access and the necessary skills can add, change or delete content on Wikipedia.

The task structure underlying the collaboration in Wikipedia is highly modular and granular, i.e. consists of various subtasks of varying size. That is, people can undertake tasks of varying sizes and types depending on their interests as well as the time and effort they are willing to spend. They can provide detailed articles on topics they care about, just correct grammar or spelling mistakes, add some references or links or draw attention to problematic content. As Yochai Benkler has argued in his analyses of large-scale peer-production of information and culture on the Web, the modularity and granularity of tasks is an essential factor for the success or failure of large-scale collaborations on a voluntary basis (Benkler 2002; Benkler 2006: 101f).

Concerning the quality control mechanisms, Wikipedia also differs from traditional encyclopedia. In Wikipedia, there is no editorial board or other forms of pre-publication

[23] For further policies confer http://en.wikipedia.org/wiki/Wikipedia:List_of_policiesas well as http://en.wikipedia.org/wiki/Wikipedia:Five_pillars [date of access: 12.02.2010]

review. Content is not vetted before being published. Instead, quality assessment in Wikipedia happens *after* publication. Since anyone can correct content, this post-publication quality control resides again in the hands of the community of all Wikipedia contributors and not in a pre-selected board. On the one hand this means that there is a multitude of watchful eyes, but on the other hand there is no guarantee that an entry has been reviewed for accuracy. Since any saved change is instantly visible it is just as well possible that at the time of reading one finds and excellent article – or just blunt misinformation.

This concept of open contribution and post-publication reviews has not been the initial model of quality assessment. Wikipedia is the successor of Nupedia, a system which was also intended as a comprehensive online encyclopedia, but had an elaborate system of peer review and allowed only scholars, i.e. qualified contributors to participate.[24] This rigid system of evaluation had the consequence that in the three years of its operation only 24 articles were produced on Nupedia. Compare this to the statistics above according to which after three years, the English-language Wikipedia alone almost hit the mark of 1.000.000 articles. Hence one initial conclusion to be drawn from this comparison between Nupedia and Wikipedia is that Wikipedia has produced much more content by involving more people.

However, from an epistemological perspective, size it not the only crucial criterion for assessing the merits of an encyclopedia, accuracy, the quality of information matters just as much. A central question therefore is whether Wikipedia provides accurate information. There have been several comparisons between Wikipedia and traditional encyclopedia. The comparison conducted by the journal *Nature* for instance has stirred a lot of debate around the comparative merits of Wikipedia and the Encyclopedia Britannica. One of the most important statements probably consist in the following quote: “The exercise revealed numerous errors in both encyclopaedias, but among 42 entries tested, the difference in accuracy was not particularly great” (Giles 2005: 900).

Nature’s methodology for the comparison consisted in sending entries either from Wikipedia or Encyclopaedia Britannica covering a broad range of scientific topics to reviewers. The reviewers were not told whether the article they received was from Wikipedia or the Encyclopaedia Britannica and had to analyze the number of major and

[24] All information on Nupedia from http://en.wikipedia.org/wiki/Nupedia [date of access: 11.02.2010]

minor mistakes in the articles. All in all 42 reviews were obtained which formed the basis for the statement above that at least when it comes to scientific content the "difference in accuracy was not particularly great" " (Giles 2005: 900). At this point I do not want to get into the details of comparing Wikipedia with the Encyclopaedia Britannica beyond stating that since both of them contain errors, one should always use any work of reference with caution. The results as well as the methodology of the study have been hotly debated not only between *Nature* and the Encyclopaedia Britannica. [25]

Nonetheless, a crucial insight to be obtained from this study concerns the fact that Wikipedia works well *on average*. That means that most articles at most points of time deliver quite accurate information. It has been argued that the older the articles are, the more mature and reliable they are (e.g Tollefsen 2009). Given the dynamic nature of Wikipedia, however, there is no guarantee that no one has inserted blunt misinformation just before I access the entry.

From a socio-epistemological perspective, Wikipedia is an interesting example of epistemic social software. And indeed, as is outlined in the second part of this thesis Wikipedia has been the social software application that has received most attention from epistemologists. The most crucial question that has been addressed concerned the validity of Wikipedia as a source of knowledge. In other words: can one trust Wikipedia to know? And if so, for what reasons? Different answers have been given to these questions (Magnus 2009, Tollefsen 2009, Wray 2009)[26] and some of Wikipedia's critics keep insisting that one should not trust it at all (Waters 2007, Keen 2008). And indeed, many of the proxies that we usually use to assess the trustworthiness of information are missing: due to its openness, people can edit who would not be considered experts on the topic. And due to the anonymity of contributions we can neither assess the competency nor the honesty of contributors. The reason for trusting Wikipedia must thus be a different one than trusting a person. Later on I argue that the reason why people trust the content of Wikipedia is that they trust the *processes* of Wikipedia. It is a form of procedural trust, not a trust in persons. Many people may not know how Wikipedia works and trust it blindly. They may simply use the information on Wikipedia, because it is the first information that appears in their Google search results.

[25] The documents illustrating the debate between the journal *Nature* and the Encyclopaedia Britannica can be retrieved here: http://www.nature.com/nature/britannica/index.html [date of access: 11.02.2010]
[26] These approaches are portrayed in the second part of this thesis in some detail.

However, if people know how Wikipedia works and trust it, then they trust it because they attribute some epistemic authority, some trustworthiness to the process by which Wikipedia generates information. They do not trust specific persons; they trust Wikipedia as a system that is based on a distinct process of information provision. This process is characterized by a multitude of users that can edit and change information immediately combined with a mechanism to undo revisions easily and quickly as well as a system for tracking these changes and making them visible. It is a system enforcing open access with minimal entry barriers combined with mechanism of making editing patterns transparent.

The rationale behind this openness of Wikipedia is *scale*. More people can provide more information faster. However, a second point is crucial: the more people scrutinize and can easily change content, the less likely it is that error and bias remains undetected. But there is no guarantee that error or bias gets detected. It is only more likely if a multitude of diverse people participate in this process. This is a probabilistic statement. Wikipedia provides a lot of information on many topics due to the multitude of volunteers. But if Wikipedia had not provided good information it would not be used to the extent it is used. On average, Wikipedia works well as a source of information. However, the problem lies exactly in the term on average. Since Wikipedia is a dynamic system, information can be changed by anyone anytime. Hence, there is no guarantee that the information provided at the moment one accesses Wikipedia is correct or a blatant lie. While Wikipedia might be trusted as a system because of certain characteristics, this overall trustworthiness does not help us to assess the trustworthiness of a specific claim in Wikipedia. This is the crux of statistical reasoning: it works well on average, but we can never be certain about a specific instantiation.

To conclude, the benefits of Wikipedia are to provide enormous amounts of information in a cost-efficient and very fast way by taking advantage of a multitude of volunteers equipped with diverse skills and knowledge. Due to this diversity not only a broad range of topics can be covered. It is also more likely that topics are portrayed from multiple perspectives, in particular if the Neutral Point of View-policy is followed. And due to certain social norms as well as technical features of the platform (e.g. the history pages), Wikipedia has a quite high level of accuracy as well. The average reliability of Wikipedia articles is quite high, but the term average also implies an important disadvantage of Wikipedia. The combination of the dynamic character of Wikipedia with the absence of a centralized control mechanisms and pre-publication quality control implies that there is no guarantee that for the quality of a specific article at any given point of time. Moreover due to the low entrance barrier and the possibility of anonymous contributions the system easily grows in scale, which has been necessary for the success of Wikipedia. However, the same criteria make Wikipedia quite vulnerable to system attacks – be they unintentional or intentional.

Users may unintentionally cause harm by deleting information by mistake or by adding content they think is correct, but actually is not. However, incompetence is not the only problem. Dishonesty and bias is a major threat to Wikipedia. In 2007, a tool called WikiScanner received quite a lot of attention and media coverage. WikiScanner is a search tool that traces IP addresses of those who change Wikipedia entries anonymously. The tool was developed by Virgil Griffith, a graduate student at CalTech. Griffith says that the inspiration for this tool has been the revelation that the offices of Congress members had been editing their own Wikipedia entries (Borland 2007). He wanted to find out whether other companies and organizations also edit entries in ways that served their interests. And indeed, by tracking the IP addresses of anonymous editors, this tool unveiled that numerous organizations were editing Wikipedia articles anonymously in a way that served their particular interests. On his website, Griffith concludes: "Overall - especially for non-controversial topics - Wikipedia seems to work. For controversial topics, Wikipedia can be made more reliable through techniques like this one".[27] I return to Wikipedia as a prime example of epistemic social software at various instances in this thesis.

2.4.2 Social Bookmarking: Sharing Information and Co-Creating Folksonomies

Another example of epistemic social software are social bookmarking services, applications to store and share links with others. By now, numerous applications exist for different types of content.[28] Some of them focus on social news, e.g. *Digg*[29] or *Reddit*[30], some on blogs, e.g. Technorati[31], some on micro-content, e.g. Twitter[32], and

[27] FAQ of WikiScanner: http://virgil.gr/31
[28] For a list of the Top 20 Most Popular Social Bookmarking Websites (February 2010), confer http://www.ebizmba.com/articles/social-bookmarking-websites [date of access: 14.02.2010]
[29] http://digg.com/ [date of access: 14.02.2010]
[30] http://www.reddit.com/ [date of access: 14.02.2010]

still others target scientific content. *Connotea*[33], for instance is a free online reference management and social bookmarking service for scientists created by Nature Publishing Group. *CiteULike*[34] is a similar system sponsored by Springer. Both systems offer additional features that are useful for scientific work, such as exportability into software tool for managing bibliographies via Bib TeX files. One of the most well-known and earliest examples of a social bookmarking system, however, is *Delicious*.[35] Hence, I have chosen *Delicious* to outline the basic characteristics of social bookmarking services.

Delicious

Delicious is a social bookmarking service that enables users to tag, save, manage and share web pages. By installing an extension into their browser (e.g. Firefox) users can save bookmarks not only on their local hard disk, but also on the *Delicious* servers. As a consequence, those bookmarks saved on *Delicious* can be accessed from any computer with Internet access. The user only has to log into the *Delicious* website at http://*Delicious*.com/. While increasing the accessibility of information and the flexibility of users, this characteristic alone would not render *Delicious* of *socio*-epistemological interest. However, there are indeed two features of *Delicious* which are of socio-epistemic relevance: sharing of links and social tagging. Users are invited to share interesting links with others by making them public or sending them via Twitter or email to others. Moreover, by *tagging* their links, i.e. by adding key words to a link for later retrieval, users also co-create a bottom-up classification system of content *while* saving their personal bookmarks. Finally tagging itself can be used as a means to share content. One the one hand users can search for content with specific tags, i.e. I can search the links of all Delicious users by using a certain tag, such as "epistemology", "Paris", or "knowledge". One the other hand, using a *common tag* can also be used to collaboratively set up shared, decentralized databases of content with colleagues or friends. In the following I explain these different socio-epistemic aspects of *Delicious* in some detail.

---

31 http://technorati.com/ [date of access: 14.02.2010]
32 http://twitter.com/ [date of access: 14.02.2010]
33 http://www.connotea.org/ [date of access: 14.02.2010]
34 http://www.citeulike.org/ [date of access: 14.02.2010]
35 http://*Delicious*.com/ [date of access: 14.02.2010]

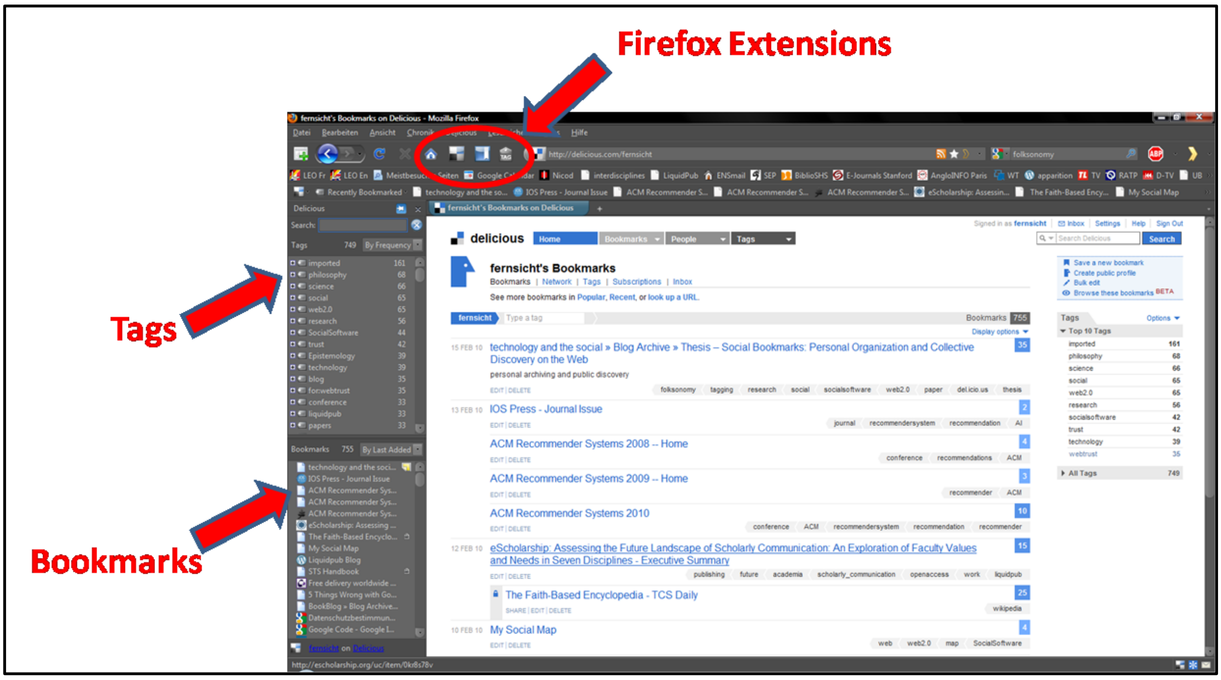

Figure 9: Delicious Embedded in a Firefox Browser, Basic Functionalities

Explicitly Sociality: Sharing content with others

Since *Delicious* is a *social* bookmarking service, the private and isolated user of this software is invited to publish his bookmarks, to make his personal selection of interesting information available and accessible to others. This can be done in different ways. On the one hand, user can decide for each link that they save whether they want to make this link public to the whole *Delicious*-community of users. Another possibility lies in sharing content just with selected other users. *Delicious* offers the possibility to share links with selected others by using twitter, email or by notifying other *Delicious* users on their *Delicious* accounts. To actively *retrieve* links saved by others, users can search the public database of links and they can use the subscriptions and network features of *Delicious*.

Hence, a crucial socio-epistemic benefit of social bookmarking systems consists in the *sharing of information*. For instance, someone interested in a certain topic, let's say philosophy of mind, can benefit from the bookmarks of someone who is an expert in the field. Instead of testing and assessing all possible websites available on this topic, he can rely on the previous assessment of someone he considers to be competent in the field. This process is similar to deference to authorities encountered in other social fields, such as science. Another option consists in using a keyword to search for links saved with this keyword. Here, one is making use of the link collections of all Delicious-users to find certain information instead of screening the links of one particular person. Nonetheless, in both cases, information is shared.

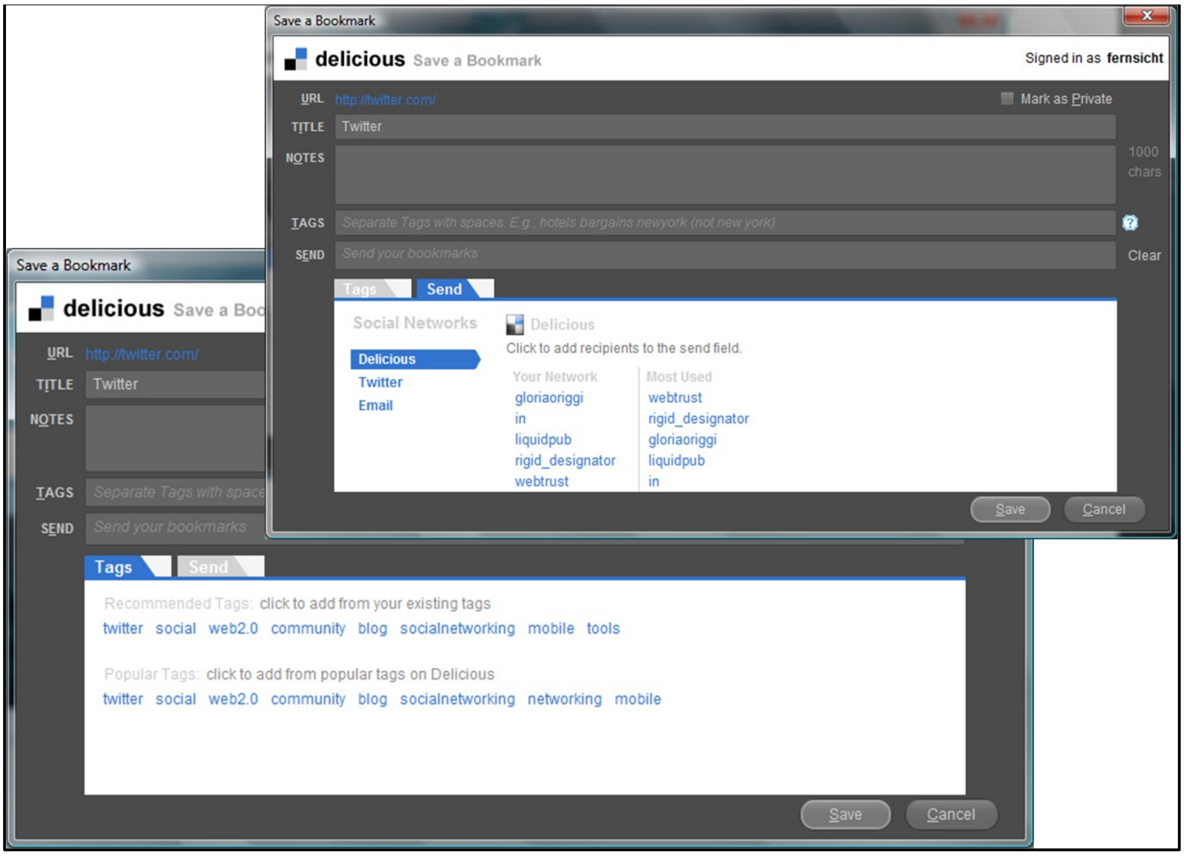

Figure 10 Delicious: Tagging and Sharing Content

Implicit Sociality: Social Tagging and the Co-Creation of Folksonomies

In addition to this invitation to *share information* with others, *Delicious* is of socio-epistemic interest for a second reason. In saving bookmarks on *Delicious*, users *co-create bottom-up classification systems,* often referred to as *folksonomies*. Luciano Floridi defines and explains folksonomies as follows:

> "A folksonomy, from 'folk' and 'taxonomy', is the aggregated result of the social practice of producing information (metainformation, to be precise) about other information (e.g. a photograph) through collaborative classification, known as social tagging (e.g. the photograph receives the tags "New York", "Winter", "Statue of Liberty"). It works bottom-up, since it is left to the single individual user or producer of the tagged target to choose what to classify, how to classify it, and what appropriate keywords to use in the classification" (Floridi 2009).

Hence folksonomies are bottom-up classification systems that emerge from a process labeled *social tagging*. In general *tagging* refers to adding key words to content of the Web. *Social tagging* then refers to systems in which multiple users tag material. There are innumerous things that can be tagged on the Web: books (Amazon.com), articles (CiteUlike.org), pictures (Flickr.com), videos (Youtube.com), blogs (Technorati.com), short text messages (Twitter.com), music (last.fm). Tagging provides a mean to classify and categorize information for later information retrieval. In contrast to top-down classifications where categories are predefined into which material can then be sorted, tagging is a bottom-up approach. It is the users who choose the tags not from a fixed set of possibilities, but creates a free-text terms that they think classifies the content to their requirements. These tags may be semantically totally unrelated to the content and may still be completely reasonable. For instance, if you wish to make certain pictures on Flickr.com available only to your friends you may add a tag of random symbols such as *&hrqait@* and tell this string to your friends. This way, they can access the pictures by using the string for their search.[36]

Folksonomies have received a lot of attention especially in information science since they deliver an antidote to ontologies as top-down classification systems designed by experts. A lot of research has been conducted on comparing the folksonomies with top-down approaches and different authors have emphasized either the advantages or the disadvantages of top-down and bottom-up approaches. For instance, while flexibility and scalability are some of the major advantages of folksonomies, problems may arise with inaccurate descriptions. Recently, in the library and information sciences, there have been attempts to combine both approaches to combine their merits (e.g. Hidderley and Rafferty 2006). Comparing different classification systems goes beyond the scope of this thesis. For the moment it is sufficient to acknowledge that folksonomies are an epistemic product that is the result of a social process. They are socially created classification systems which provide the basis for later information retrieval not only of the individual user who tagged the content, but also of other users of the system.

On *Delicious* users tag other websites, i.e. the links that they store. Whenever users save a bookmark on their *Delicious*-account, they are informed not only about how many other community members have also saved this link. If the items you are tagging are

[36] There are of course other ways to restrict access to content to certain groups of people only. The point of the example was rather to elucidate different usages of tags.

already tagged by someone else, *Delicious* will *propose* different tags for this website. More specifically, *Delicious* offers its users *recommended* as well as *popular tags* to classify their links. While the *recommended tags* are inferred from a user's own tags used to classify *other links* on *Delicious*, the *popular tags* are aggregated from the tags *other Delicious*-users have used for tagging the same Website. This means, that tags used by other users are recommended to you as possibilities to classify your own link. This example shows one way in which tagging on *Delicious* is social even if it is used for private archiving of links only. The moment people use popular tags to classify their content they make use of aggregated social information provided by other users. And even if they do not use these popular tags, their own tags will find their way back into the collaborative repository by being part of the pool of tags that may end up being recommended as popular tags to other users. Thus, even in its most solitary usage, the use of tags to store content on *Delicious* is a socio-epistemic process.

Combining Tagging and Sharing: Common Tags & Shared Databases

While it is possible to use tagging as a private means to classify information, to always refrain from using popular tags and to always click on the "mark as private"-button when saving a bookmark, *Delicious* mostly is used in more social ways. We have seen before that *Delicious* encourages the sharing of information in various ways. Not only is it possible to make one's links public and by such add them to the collaborative repository of related information, driven by personal interests and creative organization"[37], one can also send links via email or twitter-messages to friends and collaborators or save it for them on their *Delicious* account.

However, another possibility directly related to the process of tagging is to share content by using a *common tag*. I am taking part in a research project called LiquidPub. It addresses the question of how Web2.0 technologies change or should change the way scientific knowledge is produced, disseminated, evaluated, and consumed.[38] In this project we use *Delicious* as a tool to share links to interesting material related to the research topics of the project. By adding the tag "liquidpub" to links that any participant of the research project saves on her *Delicious* account and considers relevant for the project a shared de-centralized database of links is created. This database is not only

[37] http://*Delicious*.com/help/faq [date of access: 21.11.2009]
[38] For further information confer the project website at: http://project.liquidpub.org/ [date of access: 12.02.2010]

visible to the project member, but to every *Delicious*-user, who types "liquidpub" into the search field of his *Delicious*-account. Moreover, if other users, who are not part of the project, use the same tag "liquidpub" their links will also be added to the database. This may be beneficial as well as detrimental. It is beneficial if this user uses the tag "liquidpub" to classify similar content and by doing this adds new material to the virtual database. However, if he uses the same tag for a different purpose, for instance to save and share information about pubs offering good beer for reasonable prices in central London, the utility of the tag for the research group will diminish. Especially if many people start using the tag "liquidpub" for other purposes, the information related to scientific publishing will be drowned by the *noise*, i.e. information about bars in London. This example is not only an example for the social use of tagging. It also emphasizes the relevance of the *specificity* of tags.

To conclude, *Delicious* as a social bookmarking service can serve as a prime example of epistemic social software. To reach different epistemic purposes (such as sharing, generating meta-data, creating classifications, etc.) it uses and combines various social mechanisms. *Delicious* uses aggregational mechanisms to recommend tags to its users and encourages them to share information with others via various modes. And I argue later on that it is this intelligent combination of different types of epistemic sociality that is characteristic of the most successful Web2.0 applications. *Delicious* primarily is a tool for sharing existing content, such as websites, blogs, movies or music. But *en passant*, Delicious users also create new epistemic content: meta-data and folksonomies are forms of second-order information that emerge as a by-product of social tagging. The next example of epistemic social software, Innocentive.com, however, is an example of a website which focuses exclusively on the creation of new knowledge: it is an "open innovation marketplace".

2.4.3 Innocentive.com: Finding the Problem Solver

"If You Have a Problem, Ask Everyone" (Dean 2008).

In the case of Wikipedia numerous people collaborate in creating an encyclopedia by compiling *existing* knowledge. Creating *new knowledge* is explicitly not the goal of Wikipedia, indeed "No original research" is one of three core content policies of

Wikipedia.[39] Delicious, focuses on sharing existing content, but creates epistemic content in form of folksonomies. InnoCentive, by contrast, explicitly aims at eliciting the creation of new knowledge.

Innocentive.com was founded in 2001 with financial help of the pharmaceutical company Eli Lilly. The goal of this start-up was to harvest the potential of the Web for research and development (R&D). While primarily targeted at companies, Innocentive.com added a non-profit area in 2007. The idea behind Innocentive.com is quite simple. According to them there are in principle four ingredients necessary for innovation: problem seekers, problem solvers, some regulations on the interaction between problem solvers and problem seekers and a space where they can meet. Innocentive.com intended to become exactly this place were solvers and seekers meet under specified circumstances.

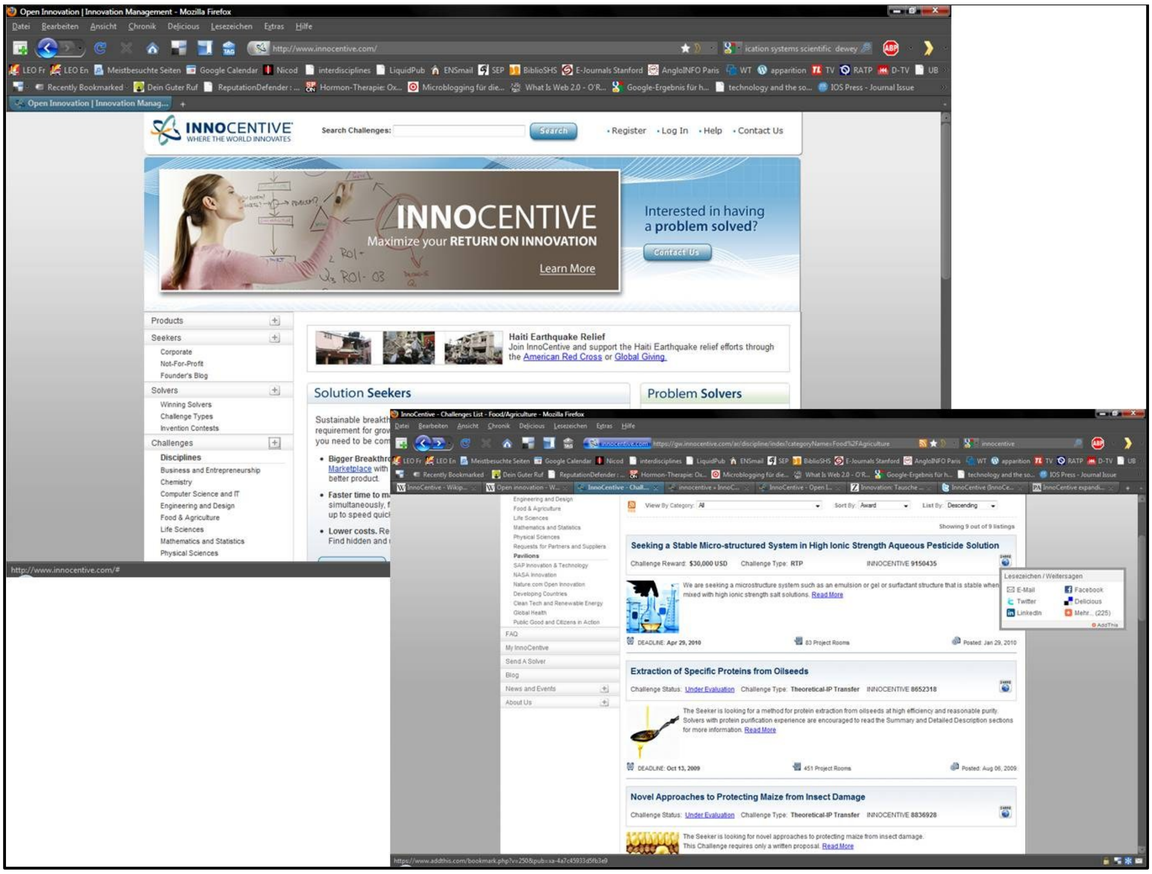

Figure 11: InnoCentive.com[40]

[39] http://en.wikipedia.org/wiki/Wikipedia:No_original_research [date of access: 15.02.2010]
[40] http://www.innocentive.com [date of access: 15.02.2010]

The problem seekers*, called InnoCentive Seekers* can be corporations, non-profit organizations as well as government bodies who offer an award for those that can solve their problem. *InnoCentive Solvers* are people who can solve these problems and will be rewarded. According to InnoCentive, there are currently 200,000 people in more than 200 countries registered at InnoCentive as potential solvers. The awards offered range from \$5,000 to \$1 million[41], while most awards range between \$10,000 and \$25,000. Whoever finds the best solution will be awarded with a financial reward. Sometimes more than one solution will be rewarded and the money will be split. By the time of writing this thesis 539 submissions have been awarded adding up to prize money of more than \$4 million.[42] The amount of money offered for a challenge differs with its complexity, the tasks involved and the criteria for successful resolution. And given the price span of \$5,000 to \$1 million these differences are indeed profound. There are four types of challenges: *InnoCentive Ideation Challenges, InnoCentive Theoretical Challenges, InnoCentive RTP (Reduction to Practice) Challenges and InnoCentive eRFPs (electronic Request-for-Partners) Challenges.*[43]

An *Ideation Challenge* is a broad question formulated to gain new ideas, i.e. it functions similarly to a brainstorming and the submissions are usually only about two written pages. Examples of such *Ideation Challenges* were to find ideas for increasing the public transportation use to reduce greenhouse gases in Chicago or ideas to improve banking processes in the developing world.

Theoretical Challenges are more complex and require a detailed solution of a problem. One example of a *Theoretical Challenge* for instance consisted in finding new methods to analyze consumer emotions, another one in developing a statistical approach to deal with the variability in blood glucose data collected during continuous glucose monitoring. For the analysis of consumer emotions only a detailed description of the solution was required, whereas for the statistical method a written proposal had to be amended by a source code.

---

[41] Up to \$1.000.000 will be rewarded for the development of biomarker for measuring disease progression of Amyotrophic Lateral Sclerosis. http://gw.Innocentive.com/ar/challenge/8305421 [date of access: 08.12.2009]

[42] http://www.Innocentive.com/crowd-sourcing-news/innocentive-at-a-glance/ [date of access: 15.02.2010]

[43] All information about the challenges and the examples were obtained from: http://www.Innocentive.com/ [date of access: 15.02.2010]

Reduction to Practice Challenges are even more detailed and the problem solver needs to provide empirical data or physical samples to support his claims. Hence, an idea for a solution is not enough. Instead the solution must be validated by the problem solver and must succeed possible replication through the problem seekers. An example is the InnoCentive Challenge 9150435 "Seeking a Stable Micro-structured System in High Ionic Strength Aqueous Pesticide Solution".[44] To achieve the reward of $30,000 USD solvers need to provide a prototype that can be tested by the seekers before they release the reward.

InnoCentive eRFPs (electronic Request-for-Partners) *Challenges* finally allow problem seekers to search for partners in solving a challenge. An example for such an eRFP Challenge is the InnoCentive Challenge 9028671 in which the problem seeker is "looking for a partner to purify an enzymatically active recombinant human protease"[45]. Here solvers need to submit a proposal and after evaluation the seekers will decide whether they want to establish a collaborative partnership.

The R&D focus of InnoCentive becomes obvious when taking a look at the disciplinary classification of challenges. Besides *business & entrepreneurship* challenges are classified into the categories of *chemistry, computer science & IT, engineering and design, food & agriculture, life sciences, mathematics and statistics* as well as *physical sciences*. Moreover, different partners offer *Pavillons* that are either named after these partners or after more general topics that they address. At the time of writing this thesis the following *Pavillons* were open: SAP Innovation & Technology, NASA Innovation, Nature.com Open Innovation, Developing Countries, Clean Tech and Renewable Energy, Global Health, Public Good and Citizens in Action.[46]

But how and why does InnoCentive work? What is its socio-epistemological relevance? According to its self-description "InnoCentive is a prime example of open innovation's crowd sourcing model, where product development costs shift from the enterprise to the market to benefit both."[47] Hence, from a financial perspective InnoCentive is successful, because it saves companies costs. Since InnoCentive is based on a "pay for performance"-model this is a very low-risk environment for companies. It is the problem

[44] https://gw.Innocentive.com/ar/challenge/9150435 [date of access: 15.02.2010]
[45] https://gw.Innocentive.com/ar/challenge/9028671 [date of access: 15.02.2010]
[46] https://gw.Innocentive.com/ar/challenge/marketPlace [date of access: 15.02.2010]
[47] www.Innocentive.com/_.../InnoCentive_Corporate_Overview.pdf [date of access: 15.02.2010]

solvers who have to advance the funds as well as their work power – and it entirely unclear whether they will earn any money from it. Such a model externalizes costs from the corporations - it shifts them from the seekers to the solvers. If that were the only market model, this would be a problematic development, especially if the preparatory effort is high. Nonetheless, since the participation is voluntary and given that economic considerations are not at the heart of this thesis I turn to the socio-epistemological relevance of InnoCentive.[48]

What are the socio-epistemological reasons for the success of InnoCentive? What type of epistemic sociality is tapped with systems like Innocentive.com? That InnoCentive is involved in *epistemic* processes is obvious, even more obvious than in the previous cases. From all Web2.0 examples provided so far, InnoCentive is most directly related to epistemic questions, because it supports the creation of new knowledge. And in its focus on research and development, on science and engineering, InnoCentive is a platform that supports epistemic processes most closely related to those prevalent in science.

But in what ways is InnoCentive *socio*-epistemic? As noted above, InnoCentive describes itself as a *crowd sourcing model of innovation* and clearly, some sort of *wisdom of the crowds* is used here to solve epistemic problems. In Part 3 of this thesis, I outline a comprehensive socio-epistemological model of epistemic social software and I return to InnoCentive in some detail. For the moment it should suffice to note that there are several characteristics of InnoCentive that are relevant from a socio-epistemological perspective because they shape the ways in which knowledge is created with InnoCentive.

First of all, InnoCentive is a *price-based* system, in the sense that the *winner takes it all*. Although there may be several winners, it is necessary to *come first* to be rewarded. Once someone else has solved the problem, those come second to go away empty-handed. This process inserts an element of *competition* into the epistemic process. Secondly, despite this focus on competition, it is in principle sufficient if *one* person or *one* team can come up with a solution to the problem posed. In that sense one may argue

[48] For the moment I also neglect the copyright and patenting issues involved in InnoCentive. Copyright and patenting issues are clearly relevant in assessing the long-term epistemic merits and dangers of systems as many legal scholars have convincingly argued (e.g. Benkler 2006, Koepsell 2003) However, they go beyond the scope of this thesis.

that InnoCentive is not really an example of *social* software, because in the end, problems can be solved by individuals: cooperation and collaboration is possible, but it is not needed. Only *InnoCentive eRFPs* are explicitly targeted at enabling and supporting collaborations. If their scopes allow for it, all other challenges can, at least in principle be solved by individuals if a single person possesses all the skills necessary for solving the challenge. Especially for the less complex problem, such as the *InnoCentive Ideation Challenges,* individuals can participate and win the rewards on their own.

However, even in this case InnoCentive is *social,* albeit social not in the sense of *collaborative*. It makes use of a pool of differently skilled people to find the one person that can solve a problem like a needle in a haystack. Taking another look at the statistics may reveal the sheer number and diversity of potential solvers that InnoCentive can tap: In February 2010 InnoCentive had more that 200.000 registered problem solvers from more than 200 countries. That the number is continuously growing becomes obvious in comparison to the data of December 2009, when they had "only" registered 160.000 solvers from 175 countries.[49]

Being able to tap such a large pool of people, means being able to solicit an incredible amount of abilities, skills and work power. Problems can be solved faster, if different people work simultaneously in different directions instead of one person trying out one alternative after the other. Moreover, the more people, the more perspectives on a problem are possible. Sometimes the initial framing of a problem may indeed hinder its solution, an effect known as *incubation* in psychological literature (e.g. Smith and Blankenship 1989). Indeed, finding someone who looks at a problem from a different angle is one of the main success factors of InnoCentive. If someone from a different background, with a different perspective approaches the same problem, it is quite unlikely that he will use the same methods. When organizations and companies open up their research problems to outsiders and offer money for this solution, they have probably already tried to solve the problem themselves and failed.[50] One of the reasons

[49] These data were available on http://www.Innocentive.com/about-us-open-innovation.php in December 2009.

[50] Of course they may also just have figured that it is cheaper to make a competition instead paying their own R&D department. Netflix, a DVD-rental service, for instance, opened up a competition for the best collaborative filtering algorithm to predict user ratings for films. They offered $1.000.000 for the best algorithm. While this sounds like a lot of money – and probably was for the team who won the prize, it is nonetheless nothing compared to Netflix's profit. For further information confer http://www.netflixprize.com/ [date of access: 15.02.2010].

for this continuous failure might be wrong starting assumptions, a wrong framing of the problem at hand. By contrast someone new, who is not polluted by this erroneous framing, might come up easily with a solution to the problem, simply by following an entirely different strategy. She might come up with a new idea, propose a different method, change different parameters and an insurmountable problem might turn out to be pretty simple. Or it might be still hard, but solvable.

The insight that *diversity* can enhance creativity, problem solving and innovation is not new. It has been discussed endlessly in literature on knowledge and innovation management, in creativity classes and it is also the rationale behind interdisciplinarity in academia and elsewhere. It has even been argued that many scientific breakthroughs have been caused by a change of perspective, but looking at things differently, by using "novel methods of representation" (Toulmin 1953: 35). One of the most interesting recent accounts of diversity is Scott Page's formal analysis concerning the power of diversity in problem solving (Page 2007). The bottom-line of his argument is that cognitive diversity trumps ability when it comes to problem solving. Looking at things from a different angle, making use of different perspectives, heuristics, interpretations and predictive models makes problem solving easier. The main conclusion from this is that one should foster and encourage diversity. Cognitive diversity, different ways of seeing, interpreting and approaching problems are often crucial for innovation, for the creation of new knowledge, for the solving of epistemic problems. I return to these issues in Part 3 of this thesis. For the moment it suffices to know that systems such as InnoCentive make use of a multitude of diverse epistemic agents, some technical infrastructure and a rewards system to create new knowledge in a highly successful way. Hence, systems such as InnoCentive have a clear socio-epistemological relevance.

2.4.4 Recommender Systems: Navigation Based on Advice

> "We are leaving the age of information and entering the age of recommendation" (Anderson 2006).

In contrast to my previous three examples, Wikipedia, Delicious and InnoCentive, recommender systems are not a specific example of epistemic social software, but rather denote a type of technology embedded in many Web2.0 technologies. Recommender systems are currently a vivid field of research and there are numerous special issues, conferences and workshops devoted to the topic.[51] Hence, giving an overview over the field in just a short section is a futile task. However, given the ubiquity and socio-epistemic relevance of recommender systems, they cannot be omitted from any socio-epistemological analysis of the Web2.0. In the following I therefore outline some major aspects of recommender systems to the extent to which they are relevant for this thesis. I return to the topic in some detail in Part 3 of this thesis, therefore this section should only be considered as a primer to outline some basics of Recommender Systems (RSs).

In defining what RSs are I turn to description offered on the website of the *AMC Conference on Recommender Systems*, one of the major conferences in the field. They describe RSs as “software applications that aim to support users in their decision-making while interacting with large information spaces. They recommend items of interest to users based on preferences they have expressed, either explicitly or implicitly.”[52] Recommender Systems are used on numerous commercial websites and are employed in various ways. The emails from Amazon.com reading “Amazon.com has new recommendations for you based on items you purchased or told us you own” - are just one well-known example. Indeed, Amazon.com is not only one of the most well-known examples, it has also been among the first Websites that has employed recommender systems quite soon after its establishment in 1995 (Schrage 2008).

Given the increasing volume and complexity of information on the Web there seems to be an increasing need for tools that support the navigation through the information space, tools that offer advice on how to chose and distinguish the important from the unimportant, on how to search and find what is needed (Resnick and Varian 1997). Indeed, one of the strengths of recommender systems may lie in *serendipity* – in finding something one was not even aware that one needed it. As Jeffrey M. O'Brien notes, “[t]he Web, they say, is leaving the era of search and entering one of discovery. What's the difference? Search is what you do when you're looking for something. Discovery is

---

[51] There have been workshops at the European Conference on Artificial Intelligence (ECAI) (http://www.eccai.org/ecai.shtml), as well as different ACM Conferences on Recommender Systems (http://recsys.acm.org/2010/). Special Issues on Recommender Systems have for instance been provided by ACM Transactions (e.g. http://tweb.acm.org/RecSysSpecialIssue.html; http://portal.acm.org/citation.cfm?id=1096737.1096738 or http://portal.acm.org/toc.cfm?id=963770), IEEE Intelligent Systems (http://www.computer.org/portal/web/csdl/abs/mags/ex/2007/03/x3toc.htm) or the International Journal of Electronic Commerce (http://portal.acm.org/toc.cfm?id=1278152). [date of access: 14.02.2010]

[52] Website of the 3rd ACM Conference on Recommender Systems: http://recsys.acm.org/2009/ [date of access: 13.2.2010]

when something wonderful that you didn't know existed, or didn't know how to ask for, finds you.” (O'brien 2006).

Types of Recommender Systems

In an state-of-the-art report on recommender systems, Adomavicius and Tuzhilin note that two of the reasons why there has been so much work on recommender systems in industry as well as in academia during the last years is that it not only is a problem-rich research area, but also a field with numerous practical applications (Adomavicius and Tuzhilin 2005). Being aware that the field has flourished since then, their basic classification of recommender systems into *content-based, collaborative filtering* and *hybrid approaches* nonetheless seems still valid. However, in addition to this classification, there is one more recent development in the field of recommender systems that is of crucial relevance for this thesis: *trust-based or trust-aware recommender systems*. Before I turn to those RSs that exploit trust relationships between users, let’s first take a brief look at the basic features of classical RSs.

As noted before, RSs suggest items to users of a system that he or she might be interested in. The problem that RSs therefore face consists in *estimating* the he utility of items for users that they have not yet assessed. Sticking to the example of Amazon.com, RSs should suggest a book to a user that she has not read yet, but will like. The utility of items is mostly represented by ratings: a much liked book is a book that has received high ratings from a user; a book that should be recommended is one that would receive high ratings had the reader read it already. RSs then extrapolate from a user’s ratings of items he knows to items he does not know. Selecting the highest-rated items from those unknown items then results in the list of recommendations.

This extrapolation can be done in various ways and numerous different algorithms have been proposed. One classic classification of different methods of RSs distinguishes content-based from collaborative filtering methods. While in content-based RSs users are recommended items similar to the ones the user preferred in the past; collaborative filtering techniques recommended items that people with similar tastes and preferences liked in the past (Adomavicius and Tuzhilin 2005, Balabanovi and Shoham 1997). Hence, content-based methods calculate *similarities between items*, whereas collaborative filtering techniques calculate *similarities between users*.

As the name states, *content-based RSs* depend on a characterisation of the content of items and of user profiles expressing interests in specific content. The users preferences can either be learned explicitly, e.g. by a questionnaire asking for favorite books, or implicitly by tracking the behaviour of the user, e.g. which books she has bought, which articles she has read, downloaded, forwarded, etc.. The users profile is then compared to the characteristics of items the user does not know yet and those items with the highest degree of similarity are then recommended to the user. Since this match is calculated based on the features of an item, RSs need a representation of these items. For textual material, this representation can be created automatically.[53] However, for non-textual material, such as video or music, this representation needs to be created manually and is therefore "expensive, time-consuming, error-prone and highly subjective" as (Massa and Bhattacharjee 2004) summarize. Besides these limitations concerning non-textual material, content-based mechanisms are characterized by two additional problems: *overspecialization* and the *new user problem* – also known as the *cold start problem*. *Overspecialization* refers to the fact that one of the main benefits is one of the main limits of content-based recommendations: the focus on similarity. Users can only get recommended things that are similar to those they have already assessed. Hence the chance of useful serendipity tends towards zero. And if similarity is the main criterion, the recommendations may also be *too close* to the originally rated items, a problem which becomes obvious when considering its application to similar news articles featuring the exact same incident. Hence, the aspect of newsworthiness gets neglected in such systems. A final problem concerns the so-called cold start problem: for new users the system has difficulties making good recommendations because the user's preferences are not yet learned by the system. However, this last problem also exists with collaborative-filtering RSs.

Collaborative filtering techniques form the core of this section not only because they are more widespread by now. The term *collaborative* also indicates that they may be of even greater socio-epistemological interest. And indeed, since collaborative filtering techniques predict the utility of items for a user based on items rated by *other users*, they exploit of social mechanism for epistemic purpose and are therefore an important example of epistemic social software. The term *filtering* in collaborative filtering RS

[53] Different methods, such as *term frequency/inverse document frequency measure* (TF-IDF), have been proposed to automatically generate key words out of textual material (Adomavicius and Tuzhilin 2005:736).

refers to the process of making predictions and as such filter information whereas the term *collaboration* refers to the fact that the filtering is based on information (ratings of items) obtained from many users. Collaborative filtering techniques do not consider the content of items, but only the ratings they have received from a community of users. Hence, they are not confined to textual content, but can just as well be applied to music, videos, pictures, etc. The basic process consists in two steps. *neighbours formation* and *ratings prediction* (Massa and Avesani 2006). That means that in order to make predictions for a user, collaborative filtering algorithms first need to find like-minded users with similar tastes as the targeted user. These like-minded users are called neighbours and the algorithm has to compute a *similarity coefficient*, i.e. a quantitative indicator of the degree of similarity between the users. In a second step the rating that a user would give to an unknown item are predicted using the ratings of her neighbours weighted by the similarity coefficient. The more similar a user is the higher is his weight in the calculation of the prediction.

As noted before, collaborative filtering techniques have the advantage of being applicable to non-textual material. Moreover, they do not suffer from the problem of overspecialization (Adomavicius and Tuzhilin 2005). However, they come with some problems of their own. First of all, collaborative filtering techniques share the cold-start problem with content-based systems: when a new user enters the system and the systems does not know anything about the user's preference no useful predictions can be made. The same problem occurs when a new item is introduced into a system: as long as it has not been evaluated by some users it cannot be recommended. Two additional problems are sparseness and vulnerability to attacks (e.g. Massa and Bhattacharjee 2004). Sparseness refers to the percentage of empty cells, the lack of overlap between users that is particularly high in newly established databases. Vulnerability to attacks refers to the question of how easy it is for malicious users to sabotage the recommender systems.

Numerous *hybrid RSs* have been proposed that try to combine the benefits from both content-based and collaborative filtering technologies while trying to minimize the technical pitfalls of both. (Adomavicius and Tuzhilin 2005: 740) distinguish four ways in which content-based and collaborative filtering techniques can be combined:

> "1. implementing collaborative and content-based methods separately and combining their predictions,

2. incorporating some content-based characteristics into a collaborative approach,

3. incorporating some collaborative characteristics into a content-based approach, and

4. constructing a general unifying model that incorporates both content-based and collaborative characteristics."54

Trust-Aware Recommender Systems

Hybrid RSs, combining content-based and collaborative filtering techniques in various ways are one possibility to counter shortcomings of each type. However, there are other possibilities as well. One line of development that is of particular interest from a socio-epistemological perspective are RSs that make use of interpersonal trust-relationships between users. Different models have been proposed on how to incorporate interpersonal trust into recommender systems. In a tutorial on "Using Social Trust for Recommender Systems"[55] delivered for the ACM Recommender Systems conference in 2009 Jennifer Golbeck portrays four trust-based algorithms: Advogato by Levien (Levien and Aiken 1998), Appleseed by Ziegler and Lausen (Ziegler and Lausen 2004), MoleTrust by Massa and Avesani (Avesani, Massa et al. 2005, Massa and Bhattacharjee 2004) and TidalTrust, Golbeck's own model (Golbeck 2006; Golbeck and Hendler 2006). In the following I focus on the works of Paolo Avesani, Bobby Bhattachcharjee and Paolo Massa to briefly explain the basic ideas behind trust-based recommender systems.

Trust-based or trust-aware recommender systems change one crucial aspect: instead of letting users *rate items*, users have to *rate other users* to bootstrap the system. This seemingly minor change indeed has a major impact on guidelines of recommender systems design. To elucidate the characteristics of trust-aware RSs, I focus on the process of *bootstrapping*, a term used to describe procedures to counter the cold start problem. When a new user enters a system the system does not "know" anything about this new user and this ignorance makes it difficult to generate appropriate recommendations for her. To counteract this problem, traditionally new users were asked to rate a few items so that the system can "learn" something about the user in order to provide personalized information on interesting items for her. However,

[54] For examples and further information on content-based, collaborative filtering and hybrid recommender systems confer (Adomavicius and Tuzhilin 2005).

[55] |Golbeck's slides are available at http://recsys.acm.org/2009/tutorial1.pdf [date of access: 14.02.2010]

especially in large databases necessary correlations are scarce and thus, this procedure often turns out to be quite ineffective. In consequence, Paolo Massa and Bobby Bhattachasjee developed an algorithm for "Trust-aware Recommender Systems" and argue that the before mentioned problems "can effectively be solved by incorporating a notion of trust between users into the base CF [collaborative filtering] system" (Massa and Bhattacharjee 2004). The difference between traditional RSs and trust-aware RSs is quite simple: "[w]hile traditional RSs exploit only ratings provided by users about items, Trust-aware Recommender Systems let the user express also trust statements, i.e. their subjective opinions about the usefulness of other users" (Massa and Avesani 2006). This seemingly minor change proves to be highly effective to remedy the shortcomings of traditional RSs especially with respect to the cold start problem because "it is able to exploit trust propagation over the trust network by means of a trust metric" (Massa and Avesani 2006).

From a socio-epistemological perspective, this shift from rating items to rating users is highly significant: instead of rating content directly, users use evaluative social information about other users as a filter to predict the quality of content. In other words, *content information* (the quality or interestingness of an item) is predicted by making use of *social information* about the competency of other epistemic agents. I argue that this entanglement of social and content information is a crucial characteristic of many Web2.0 applications, an insight that is instrumental for the development of my socio-epistemological framework in Part 3.

2.5 Conclusions

In this chapter I have very briefly introduced the empirical basis of this thesis: *epistemic social software.* I have depicted different examples of such epistemic social software in some detail: Wikipedia, Delicious, InnoCentive and Recommender Systems. The commonality between the different systems portrayed is that in each case a *multitude of epistemic agents is involved in epistemic tasks*: users create an encyclopedia or classifications systems, they search for information, share links or references, they even create scientific knowledge. Each system exploits *epistemic sociality*; it utilizes a multitude of users for epistemic purposes and goals. But as I argue in Part 3 in more detail, the systems exploit *different* types of epistemic sociality.

This introduction was necessary to give an overview over the types of applications addressed in this thesis, to introduce the empirical occurrences that have stirred the analyses which are depicted in the next chapters. However, this introduction has been rather cursory. I have remained rather descriptive and only hinted at the socio-epistemological relevance of epistemic social software. There are the following reasons for this proceeding. First of all, to apprehend epistemic social software in a form that goes beyond a mere description of functionalities, a theoretical framework is needed. I have proposed to use social epistemology as a frame of reference. Hence, I need to introduce the field of social epistemology before being able to provide a socio-epistemological analysis of epistemic social software. This task is fulfilled in the second part of this thesis, where I outline major theories within social epistemology and assess their utility for the assessment of epistemic social software. However, one result of these analyses can be forestalled: so far social epistemology has only insufficiently addressed the role of technologies for socio-epistemic practices. Hence, to analyze epistemic social software, I need to augment my theoretical framework with insights from a different field to shed some light on the interplay between the social, the technical and the epistemic, which is so characteristic for epistemic social software. The field that appeared to be most appropriate to this respect are Science and Technology Studies (STS). Hence, before turning to social epistemology, I make a brief detour through the field of STS. Insights from both STS and social epistemology are fed into the development of my own socio-epistemological model for analyzing epistemic social software in Part 3.

3 Epistemic Social Software as Socio-Technical Epistemic Systems

In the previous chapter I have introduced various examples of social software and outlined my focus on *epistemic social software*. Consistent with ordinary language usage, I have often talked about *software, tools* or *applications* when referring to Wikipedia, InnoCentive, social bookmarking services or recommender systems. In this chapter, I want to correct my own terminology. I argue that instead of tools, applications or software, the examples I have depicted should better be understood as *socio-technical epistemic systems consisting of multiple human and non-human agents who interact for various epistemic purposes*. To support such an understanding of social software as socio-technical system comprising of entangled, heterogeneous agents I take a brief detour through the field of Science and Technology Studies (STS). After outlining the field and depicting some of the main approaches I draw some conclusions on the status of epistemic social software. The insights I take from STS to characterize epistemic social software as socio-technical epistemic systems are clustered around the following topics: the *entanglement of the social, the technical and the epistemic*; the *relationship between human and non-human actors*; a *performative understanding of epistemic practices and systems*; the *relationship between values and technologies.* In particular, I argue that *epistemic social software* is a term to denote systems in which social, technical and epistemic aspects are deeply entangled. These systems consist in networks of human and non-human agents. Both human and non-human agents are of equal importance for the functioning of the systems, but they differ with respect to the question of intentionality and reflexivity. In other words, only human agents can be made *responsible and accountable* for their actions and hence a difference between human and non-human agents needs to be made. During their development of socio-technical epistemic systems different values are necessarily inscribed into them. Such a *valuing of systems* in unavoidable, because each decision in developing systems in based on certain epistemic and societal background assumptions. Despite their unavoidability, such processes need to be made transparent and monitored because such embedded values can retroact on societal values in general by either reinforcing or degrading them. Finally, socio-technical epistemic systems are not closed systems. Instead they develop and change through practices. Despite the relevance of programming decisions, users can to varying degrees appropriate and change systems through their own practices. From a normative perspective, these insights imply that although to varying degrees, designers and developers as well as users need to be made responsible and accountable for the decision they make and the actions that take.

3.1 STS: A Very Brief Overview

STS is used as the abbreviation for different terms. It sometimes stands for "Science and Technology Studies", but is also used for "Science, Technology and Society", or "Social Studies of Science and Technology" (Van House 2003). In the following I do not differentiate between these notions, since it is the abbreviation STS which is most commonly used to describe the field of research analyzing the intersection between science, technology and society.

The roots of STS lie in various fields, such as history, philosophy and sociology of science and technology, anthropology, cultural studies, critical theory, feminist theory and gender studies, with each discipline approaching the field of science and technology in its relation to society with its own methodological and theoretical toolkits. STS is often considered to be a part of the more general term *science studies*, which focus on the analysis of scientific knowledge, i.e. theories, methods, evidence, etc *as social phenomena*. In doing this, science studies put emphasis on *practices and artifacts* as opposed to a *history of ideas* or reflection of *theoretical concepts* as characteristic of many approaches in philosophy of science. Moreover, science studies primarily focuses on *scientific content* as opposed to the *institutions, processes, the norms and participants* in science, which were of central interest for the *sociology of science* as exemplified by the works of Karl Mannheim (Mannheim 1936) or Robert K. Merton (Merton 1973).

In an extensive review about the state of the art of STS and its relevance for information studies, Nancy van House states that STS's "primary concern is the mutual constitution of the technical and the social" (Van House 2003: 4). She argues that STS might be instructive for information studies in understanding two major issues. One the one hand, STS according to Van House can be a "[...] source of generative understandings of knowledge and knowledge communities, processes, practices, artifacts, and machineries", in which knowledge is considered to be "[...] situated, social, shared, multiple, distributed, and embodied" (Van House 2003: 70). On the other hand, STS is useful in understanding information systems and technology as "[...] socio-technical systems, ensembles of materials, machines, people (users, designers, operators, contributors, and others), practices, representations, understandings, categorizations, and other components , interacting with and mutually constituted by one another" (Van House 2003: 72).

These quotes should be sufficient to explain my interest into STS given the topics of this thesis. Considering the goal of this thesis to analyze the epistemic relevance of epistemic social software any theory that stresses the relationship between the social, the technical and the epistemic is clearly of high relevance. And since van House connects STS even directly to information and communication technologies, her review has served as a great starting point for my explorations concerning the relevance of STS for the questions of my thesis. However, in my portrayal I go beyond the approaches introduced by van House. Hence, in the following I briefly portray the major approaches in the field of STS as well as two extrapolations with crucial relevance for my analyses: feminist STS as well as research on Values in Design (VID).

In this section I do not aim at providing any comprehensive introduction to the field of STS. Rather I focus on crucial insights concerning the relationship between the social, the technical and the epistemic to the extent to which they are relevant for the question of my thesis. For those interested in more in-depth portrayals of STS and related fields, I can only link to some central publications. A good starting point for any analysis on the intersection between the social, the technical and the epistemic is "The Science Studies Reader" edited by Mario Biagioli (Biaglioli 1998), which comprises numerous important articles in the broader field. Two seminal edited books, which mark the birth of STS as a field, are Bijker, Hughes et al. (Bijker, Hughes et al. 1987) and Mackenyie and Judy Wajcman (Mackenzie and Wajcman 1999). Specifically for Actor-Network-Theory (ANT), please confer Latour's introduction to Actor-Network-Theory "Reassembling the Social" (Latour 2005) as well as (Law and Hassard 1999). John Law also provides an extensive amount of information on ANT on the Web[http://www.lancs.ac.uk/fass/centres/css/ant/antres.htm]. Ilyes has provided an extensive overview of the current state of the art of the international STS debate (Ilyes 2006). For possible applications to information studies the ARIST chapter by Nancy von House is invaluable (Van House 2003).

3.1.1 Social Shaping of Technology (SST)

One of the first approaches to assess the relationship between society and technology within what later was construed as STS was labeled *Social Shaping of Technology* (SST). As seemingly is the case for all abbreviations and labels in STS, there are different understandings and interpretations of the label SST. Some authors conceive SST as an umbrella term for all social studies of science and technology, which includes also the Social Construction of Technology (SCOT) and Actor-Network Theory (ANT) (Williams and Edge 1996). Others clearly separate SST from SCOT and ANT and depict it as a distinct approach (Van House 2003). In the following, I continue using *STS* as the umbrella term for studies on the relationship between science, technology and society. *SST* is used as the abbreviation for "Social Shaping of Technology" as a distinct approach which was instantiated with MacKenzie and Wajcman's book "The Social Shaping of Technology: How the Refrigerator got its hum" (Mackenzie and Wajcman 1999). In the following I briefly portray the some main characteristics of SST and draw some conclusions concerning the topics of this thesis. I then introduce SCOT and ANT in the next two sections, assuming that they can best be regarded as extrapolations or critical responses to central claims in SST.

SST as introduced by MacKenzie and Wajcman (Mackenzie and Wajcman 1999) set out to undermine two prevalent assumptions about technology: the *value neutrality of technological artifacts* and *technological determinism*. A classical example of such assumed value-neutrality of technology would be to argue that it is not weapons that kill people, but those who use the weapons. Accordingly, it is argued, the weapon itself is value-neutral or value-free - it is only the usage of technology, which can be ethically assessed.[56]

Technological determinism usually refers to two claims. First, it describes the conviction that technology develops from forces within technology, that it follows an inner-technological teleology. Such a view has been predominant not only in sociological analyses of technology, but also in the field of technology assessment as well as in popular literature on technology for a long time (Wajcman 2002, cited from Van House 2003). Secondly, if such technological determinism is embraced, any social analysis of technology can only be an analysis of the *impact of technology on society*. Since

[56] I return to the issue of values in technologies at the end of this chapter in more detail.

technology unfolds according to its own inner rationale, society simply has to deal with these technological changes and the role of any analysis of technology is to prepare society do deal with those changes in the best possible ways.[57]

Hansen and Clausen argue that SST was a response to such "[...] techno-economic deterministic understandings of the relations between technology and society" and that instead of buying into economic or technological determinism SST stressed the "[...] social choices involved in the co-evolution of technology and society" (Hansen and Clausen 2003: 431). Thus, instead of considering technology development and technological change as of being driven by technology itself, SST views technology as the outcome of social negotiations between a diversity of different agents and stakeholders. As Williams and Edge argue, "SST studies show that technology does not develop according to an inner technical logic but is instead a social product, patterned by the conditions of its creation and use. Every stage in the generation and implementation of new technologies involves a set of choices between different technical options." (Williams and Edge 1996: 366). In denying technological determinism, SST also tried to change the exclusive focus on the *social impact of technology*. Instead, by disclosing the social forces which played a role in *technology development*, SST shows, that it is possible to *change* the course of technology through negotiations instead of fatalistically just dealing with consequences of uncontrollable technological progress.

Despite many criticisms around early forms of SST, which have been discussed in the last about three decades, there are four major points, I consider important to understand the relationship between the technical and the social. First of all, SST has to be applauded for disclosing and criticizing technological determinism and for remedying the exclusive focus on the impact of technology onto society by highlighting negotiations that take place during the development of technological artifacts. Thus, society is not only understood as something to be influenced by technology, but something that has an impact onto technology as well. In the course of technology

[57] Technological determinism still is quite prevalent in mainstream technology assessment nowadays. Indeed, the German translation of technology assessment usually is "Technikfolgenabschaetzung", i.e. assessment of the consequences of technology. Within the TA-discourse, alternative such as deterministic visions have been proposed. Constructive technology assessment (cTA), for instance focuses on the design process and the social choices and negotiations involved therein and is closer to SST. Such approaches, which distance themselves from simple impact-models have been more dominant in several Northern European countries and are characterized by a stronger focus in mutual shaping of technology and society as opposed to an impact of technology on society (Hansen and Clausen 2003, Cronberg 1996, Schot and Rip 1997)

design and development, decisions have to be made at various stages. For instance, when developing recommender systems, different algorithms are possible and different algorithms have different effects. Hence, putting emphasis on the different societal values that are embedded into technologies and the interactions that take place in the course of their development surely is a crucial insight to be obtained from SST.

However, SST, especially in its early forms was often accused of just having replaced one form of determinism -technological determinism - by another one, namely social or political determinism.[58] In their effort to show that technologies do not develop according to some inner-technological trajectories, proponents of SST have neglected if not negated any dynamics of technology development from within or some form of *technological momentum* (Hughes 1983). However, certain technological changes can indeed trigger other technological developments, even if this process might be mediated by economic or political forces. Thus, analyses of technology development should always take into account that what gets developed, by which means and to which ends always depends on a complicated arrangement of social, political, economic and technological forces instead of being purely technologically or socially determined.

A second, related criticism concerns SST's almost exclusive focus on the process of technology development as opposed to its usage. This view overemphasized the role developers and designers play in deciding upon the usage of artifacts and tools, while neglecting the possibility of re-interpretation and appropriation through users. It is especially this issue of *interpretative flexibility* that is a central point of critique brought forward by proponents of the social construction of technology (SCOT) (Pinch and Bijker 1984). However, I think that SST's focus on the process of technology development and the role of designers and developers remains crucial, even if appropriation and re-interpretation is possible and an important topic for analysis as well. After all, technologies come with affordances and limits of use. Thus, even though you might use a gun as a hammer or a club, firearms undeniably have changed the terms of lethal combat (cf. Kling 1992, quoted after Williams and Edge 1996).

[58] Langdon Winner's statement that "[w]hat matters is not the technology itself, but the social or economic system in which it is embedded" (Winner 1980: 122), is often used as an example of such social determinism. I return to the issue of social determinism in SST, when depicting Winner's famous – or infamous bridge-example below.

This issue of affordances and the possibility of unintended usage is also crucial for my analyses of different software algorithms later on. Although I am not denying the possibility of appropriation, this process not only becomes harder the more *hidden* technological decisions are (as often is the case for algorithms). Moreover, overemphasizing the possibility of appropriation takes away the responsibility from technology designers and software developers to think about the possible consequences their tools may have for different social groups. And this impact might indeed differ significantly. For instance, while it might be easy for someone who is technologically literate to bypass defaults set by information system developers, other users who are less experienced are much less likely able to subvert the software and hence are much more affected by decisions made by others. These users may not even be aware that there are different ways of writing a search algorithm, leading to different results and that some programmer – or a team of them, must have made a decision about the algorithms they employ.

Hence, one of the major lessons to learn from SST is that technologies are not neutral in the sense that all depends only on usage. Tools and artifacts come with affordances and limits and this is where the developers play a crucial role. In the development process, developers intentionally or unintentionally inscribe their epistemic as well as other societal values into the systems. And those values have impacts on the users, for instance by filtering information for users. Even if tools can be used in completely differently ways than intended by their developers, it would surely go too far to assume that engineering decision are without consequences.

I return to these thoughts below in the section devoted to *Values in Design*. For the moment it suffices to note that if all was a matter of usage and programming decisions were irrelevant, there would be no use in providing normative standards for the development of software. And in this extreme sense, talking about *appropriation* and *interpretative flexibility* certainly does not make sense and leads to serious problems for normative approaches. Thus, although other approaches rightly insist on the role of the users (e.g. Oudshoorn and Pinch 2005a) and on the possibility of appropriation and even guerilla tactics, SST might remind us that programming decision and thus the role of the developers are crucial for pre-selection, for making choices, for drawing the boundaries, for defining the limits and affordances of technologies.

This leads us to a third important aspect of SST. Opening up the black-box of technology development by showing that choices were made and that different social agents played a role in this process, opens up the possibility that different technologies could have been developed if other decisions had been made. And this also means that for future technological developments different directions are possible depending on who is in the position to decide upon them.

Finally, SST has been relevant in its attempt to overcome a *dichotomous understanding of technology and society*, an insight also crucial for my own analyses. Social and technological elements are understood as intertwined in their development and usage. However, this dissolution of the technical-social dichotomy is not unique to SST, but also characteristic of other STS approaches. While Bijker refers to *socio-technical ensembles* (Bijker 1995 242), John Law uses the term *heterogenous networks* (Law 1992/2001), and the *actor-network* consisting of multiple human and non-human actants is the main unit of analysis in ANT (Latour 2005).

3.1.2 Social Construction of Technology (SCOT)

The "Social Construction of Technology" (SCOT) approach can be regarded as an extension of Social Shaping of Technology as portrayed in the last section. Thus SCOT shares many of the claims of SST as well as much of the received criticism. Together with Actor-Network-Theory (ANT), SCOT is one of the two most prolific successors of SST. SCOT shares SST's critique of a linear model of technology development, which starts from basic research and then sequentially continues through applied research, development and commercialisation. Furthermore, it also stressed the processes of negotiations that take place between various social agents in the process of technology development.

As indicated by its name, SCOT is a social constructivist theory about technology development. It was introduced by Trevor J. Pinch and Wiebe E. Bijker in their seminal paper "The Social Construction of Facts and Artifacts: Or How the Sociology of Science and the Sociology of Technology Might Benefit Each Other" (Pinch and Bijker 1987). In this paper, the authors show their proximity to the sociology of scientific knowledge (SSK), especially the *Strong Programme* of the Edinburgh School as well as Collin's (Collins 1981a) programme of empirical relativism. In effect, they are trying to make the claims and insights of the sociology of scientific knowledge fruitful for the analyses of technology.

One of the central claims of the *Strong Programme* is the *symmetry principle*, which holds that successful theories have to be accounted for socially just as much as unsuccessful scientific theories. This view has stirred a lot of debate within philosophy of science, because it contravenes the quite prevalent idea, that social factors are only relevant in the case of error. That is, social factors can only be made responsible if wrong scientific theories have been mistaken to be true or if true theories have been ignored by the scientific community for a considerable period of time. From such a point of view, social factors are understood as biases that hinder rational processes in science and should better be overcome.[59] SCOT applies the symmetry principle to technology and argues that successful innovations cannot be explained by assuming that they work better than failed innovations. Rather failure and success depend on the social context and on whether an innovation was *promoted* successfully or not (Pinch and Bijker 1987; Bijker and Law 1992b).

In the same article, Pinch and Bijker (Pinch and Bijker 1987) also introduce the key concepts of SCOT: *relevant social groups, interpretative flexibility* and *closure.* Any SCOT analysis starts off with identifying different relevant social groups that hold stakes in the development of a technological artifact. These social groups are characterized by the fact that all their members "[...] share the same set of meanings, attached to a specific artifact" (Pinch and Bijker 1987 30). Between different relevant social groups, the meaning of an artifact might differ profoundly, which is accounted for by the concept of *interpretative flexibility*. Any technological artifact is developed while members of these different social groups negotiate the form and meaning of the artifact. The process by which this phase of negotiation comes to an end is labeled *closure* or *stabilization*.

This closure can be compared to the process of *consensus formation in science*. However, while in science it would be sufficient that the community of scientists agrees upon something, closure in technology depends upon the stabilization of an artifact amongst *different groups*. This stabilization is brought about by different mechanisms,

[59] Helen Longino has analyzed these debates in the first chapter of her book "The Fate of Knowledge" (Longino 2002c). The book as well as her attempt tp dissolve this rational-social dichotnomy is portrayed in detail in Chapter 4.

which include rhetorical maneuvers just as much as redefinition of the problem. In both cases, the problem neither needs to be solved, nor do other groups have to be convinced for the artifact to be closed (Pinch and Bijker 1987). The result of different stabilization procedures is the same: After a while, the negotiations that have taken place during the construction period of technologies become *black-boxed* and the artifact seems to be *given*. Alternatives, which might have existed in the course of the development, become invisible – and eventually even unthinkable. Thus, by showing that negotiations and different perspectives on the form, meaning and potential usage of an artifact have shaped its final form, SCOT is helping to open up the black boxes of technology development and innovation. This insight that things could have developed quite differently raises the awareness about the relevance of social negotiations and the role of different stakeholder in technology development.

Several insights from SCOT are relevant for the analysis of epistemic social software which lies at the heart of this thesis. First of all, by opening up the black boxes, which have been closed around existing technologies, SCOT stresses the fact that technology design and development involve numerous stages of decision making in which multiple agents are involved. This insight defeats simple forms of technological determinism. Even if their critical stance does not go far enough and SCOT avoids drawing any normative conclusions from the roles different agents play in this process, SCOT has to be applauded for uncovering the role of different social actors in the development process. SCOT also helped to understand that the same technological artifact might have completely different meaning for different social groups, by introducing the concept of interpretative flexibility. Moreover, this opening of the black boxes of technology development also shows, that technological artifacts are not just a given, but have been shaped by different social agents – and that they could look very differently had they been developed by other agents.

However, as was the case for the *Social Shaping of Technology* (SST), SCOT also has received a lot of criticism. And since both approaches share many claims their criticism is also quite similar. This is particularly true for early forms on SCOT, while later reformulations tried to account for various points of critique. One point of critique that has already been raised in the assessment of SST concerns SCOT's focus on the development of artifacts rather than their later usage and possible appropriation. SCOT as SST was also accused of social determinism and of not taking into account technological dynamics and material conditions. Moreover, it was argued that SCOT considers society as apart from technology and rather as the environment in which technology develops as opposed to ideas concerning the mutual constitution of the social and the technical (Van House 2003).

Other major problems which are targeted more directly at SCOT are related its core concepts. The first problem concerns the concept of *relevant social groups*. By focusing on *relevant* social groups in the development process, SCOT almost by definition only takes into account, those social groups who did take part in the developmental process while those, who were left out – intentionally or unintentionally – are ignored. This neglect was especially criticized by feminist scholars, who argued that this view structurally discriminates against women, who are mostly underrepresented in the developmental process and the construction phases of technologies (cf. Bath 2009). More generally, taking a look at what or who is left out, can often be at least as instructive as taking a look at the participants, as was later also recognized by proponents of SCOT themselves (e.g. Wyatt 2005; Oudshoorn and Pinch 2005a; Oudshoorn and Pinch 2005b). The notions of *closure* or *stabilization* also received criticism for being overly rigid and not accounting for the fact that artifacts can be appropriated by their users and might continue to develop after their release to the market (Van House 2003).

Another central criticism, which was especially raised by Langdon Winner (Winner 1993) concerns SCOT's neglect of power relations in technology development and the voluntary omission of normative concerns. This refusal to become normative seems to be more of a general problem of purely descriptive social constructivist approaches and it poses serious problems for critical approaches in technology design. Thus, while it is acknowledged that social groups play a role in technology development, the question which groups have how much power is just as much ignored as the question of *just or optimal distribution of power* over different actors and stakeholders. Put differently while SCOT shows the crucial role that different stakeholders play in the design and development of technologies, they refrain from making normative claims about how such a process should ideally look like. How much power should be reserved for whom? These normative issues are not addressed within SCOT and had to be tackled by mostly feminist scholars in STS (cf. Bath 2009) as well as those working in the field of values in design (Flanagan, Howe et al. 2008). And drawing normative conclusions from the insights obtained from analyses of technology development and usage is also crucial for this thesis.

Moreover, as a result of the exclusive focus on the development of technologies, the *impact,* i.e. the societal *consequences* of technologies have been mostly ignored within SCOT as originally conceived. While SCOT analyzes the relevance of the social for the technological by focusing on the role of social negotiations in the development of technology, they ignore the question of how these technologies retroact on societal and cultural values. This blind spot might be explained by SST and SCOT's initial attempt to provide an alternative to the prevalent impact models of technology assessment. However, a total neglect of the question of impact is surely not justifiable, especially since different decisions leading to different technological artifacts might indeed have different social, ethical and epistemological consequences for different agents and social groups.

3.1.3 Actor-Network Theory (ANT)

One of the most successful and influential developments in STS is Actor-Network-Theory (ANT). Even though the word "theory" occurs in its title, ANT is not a uniform theory, but rather a method (Latour 1999), a range of practices (Law 1992/2001) or an "inspirational framework [rather] than a constraining theoretical system"(Callon 2004: 65). It tries to provide a way of approaching the relationship between the technological and the social – or rather the socio-technical - that does not fall victim to the different forms of social or technological determinisms displayed in or characterized by the approaches before. ANT was developed from the 1980ies onwards by Michael Callon, Bruno Latour, John Law and others. Despite originating in the same background as other STS approaches and using similar methodologies ANT departs strongly from the accounts portrayed before by considering any "[...] social explanation of scientific facts and technological artifacts [to be] a dead end" (Callon 2004: 62). Instead of showing the impact of society on technology – or vice versa, ANT sets on to criticize essentialist concepts such as "society" altogether. The proponents of ANT argue that society instead of being an agent that can exert influence or can be influenced is an "ongoing achievement"(Callon 2004: 62). Similarly, by stating that "[...] social structure is not a noun, but a verb", Law (Law 1992/2001) stresses the *performative* character of what are considered to be core concepts in the social sciences. That these core concepts exist is the result of the *blackboxing* of the actual processes taking place. And one of the major goals of ANT is to open up these black boxes (Latour 1987). According to Callon it is exactly by "[...] jettisoning the idea of a society defined *a priori*, and replacing it by sociotechnical networks that ANT avoids the choice between sociological reductionism, on the one hand and positing a great divide between techniques and societies, on the other" (Callon 2004: 63, italics in the original).

As most STS approaches, ANT is following a strictly empirical, ethno-methodological approach that aims at learning from actors by following them through the actor-network (Latour 1999). The roots of ANT lie within laboratory studies in which social scientists with the help of ethnographic methods analyzes the creation and development of scientific fact and technological artifacts in the making (e.g. Knorr-Cetina 1981, Latour and Woolgar 1986, Latour 1987).

The main ontological unit and object of analysis is the actor-network. The actor-network is conceived as a heterogeneous ensemble of different agents, of human and non-human *actants*[60], which take part in some collective endeavor. ANT analyses how these different nodes of the network are held together temporarily, how they change over time, which translations they undergo. In doing this it focuses on the activities and processes by which power relations in the actor-network are created, stabilized and changed and how different relations within the actor-network are being transformed throughout time. Power is a *network effect* rather than an external factor exerting influence.

There are several key concepts by which these processes are characterized. One central concept is that of *translation*. Since actor-networks consist of such heterogeneous elements, multiple and continuous translation processes are taking place between different elements of the actor-network. This task of translating is conducted by various *intermediaries*, which can be human or non-human. The most obvious translation process might concern translation processes between different human agents, who participate in the collective endeavor, but who might have diverging interests and understanding of this collective activity. However, translation does also occur between human and non-human actants as well as between non-human actants. The central notion to describe these processes is *inscription*. Inscriptions can be anything from

[60] Actants are human and non-human actors which act in relation to others in an actor-network. The term was borrowed from semiotics (Callon 2004) and is meant to denote the active role of non-human agents.

photos, notes, diagrams, text, PET-Scans, etc., and they are constantly being transformed during the research and development processes (Callon 2004).

However, these transformations, the negotiations and adjustments, the different contributions of different actors become *black-boxed* after the *artifact is closed* or after a *fact is consented upon*. Thus, one similarity between ANT and the previous accounts concerns the notion of black-boxing. Just as much as methods or claims that have been consented on in science become black-boxed, transformations and adjustments are becoming black-boxed in technology. There are two instances however, in which the black boxes *are* open: One occasion is during the construction and development phase itself. The other is, when catastrophes occur, or when things are not working the way they are supposed to work (Bowker and Star 1999,Bijker and Law 1992a). Thus it is exactly these two instances, the developments as well as the occurrence of failures that ANT has particularly focused on in their attempt to show what is in the black boxes of science and technology (Latour 1987, Akrich 1992, Akrich and Latour 1992).

Of particular interest for this thesis is the concept of knowledge employed in ANT, since it is quite distinct from concepts of knowledge in social epistemology. Instead of treating knowledge as propositional or mental, ANT considers knowledge – or rather *information* - to exist only in materialized form (Latour 2005: 221ff). This focus on inscriptions and materialized forms of knowledge as opposed to a focus on "the scientists thinking hard in their offices" (Latour 1987: 237) has been captured by Latour's concept *immutable mobiles* (Latour 1987). Immutable mobiles are the inscriptions which "make it possible to record, combine, compare, summarize, link, and manipulate work performed in a variety of places to create new inscriptions and understandings out of existing ones and coordinate work across space and time (Van House 2003: 14). Knowledge in this sense is either materialized in physical objects, such as scientific articles, or embodied in skills scientists possess. While the latter instance of knowledge corresponds to what is labeled *implicit knowledge* in other fields (Polanyi 1985, Nonaka and Toyama 2003), the former would be an externalized form of *explicit knowledge*. This duality of knowledge explains the dictum "Follow the actors! Follow the text!" (Spinuzzi 2003).

Such a view on knowledge and the focus on its materiality is quite uncommon and often in sharp contrast with philosophical understandings of knowledge. In mainstream epistemology, especially in its analytic form, as is shown later on, knowledge is considered to be primarily or only propositional and often defined as *justified true belief*, with lots of discussions around whether these three terms are sufficient or all necessary.[61] Thus, ANT, just as much as analytic epistemology holds an exclusionary view on what knowledge is. Both seem to consider their views on knowledge, either as embodied and materialized or as propositional, to be the only relevant forms of knowledge. However, as Helen Longino (Longino 2002c) has pointed out, at least three different senses of knowledge exist and get frequently mixed up in academic debates. And at times this leads not only to a lot of confusion, but also to fruitless discussion and hostile misrepresentations between proponents of different approaches. *Knowledge as content*, she argues exists in materialized form. It has to be differentiated from *knowledge as knowledge production*, i.e. the practices of constructing, assessing and evaluating knowledge claims and from *knowledge as knowing, i.e.* as a state of a person towards and object which relates to philosophical debates around knowledge as justified true belief (Longino 2002c: 77ff).[62] In this sense, the concept of knowledge employed in ANT is primarily related to Longino's *knowledge as content*. And while *knowledge as knowledge-production* also plays a role in the recent more *performative turn in ANT* (Law 2004), *knowledge as cognitive agency* gets largely ignored.

This focus on knowledge in its materialized form, however, serves a certain purpose within the analytic framework of ANT. ANT considers text and the production of papers to be the primary goal of science, as opposed to maybe the attainment or approximation of truth, which many philosophers of science would assume to be a primary goal of science.[63] Text as embodied knowledge and scientific knowledge objects as "immutable and combinable mobiles" play a crucial role not only for the accumulation and distribution of knowledge, but also for enrolling allies (Latour 1987: 227).

Enrollment is another key concept in ANT. From its beginnings, ANT was concerned with the ways in which consensus and agreement on scientific facts is achieved in science. For this process of consensus formation, scientists had to convince their peers of their approaches, methods and results. And publications play a crucial role in this

[61] Many philosophers acknowledge that other types of knowledge may exist, yet frequently these are not considered central for philosophical analyses (e.g. Baumann 2006).
[62] Longino's approach is portrayed in detail in the second part of this thesis.
[63] The utility of truth as a concept of appraisal in science has however been challenged even within philosophy of science. Please confer (Longino 2002c, Solomon 2001), as well as the section on truth in Chapter 6).

process of winning allies for one's position. However, the process of enrollment does not only target scientists and other human agents (e.g. policy maker, funding agencies, etc.), but is also applied to non-human actants. The most famous example would be Michael Callon's study on the domestication of the sea scallops (Callon 1986). In this process, not only other researcher or the fisherman had to be enrolled, but also the scallops.

This example leads us to the most radical and controversial claim of ANT: the principle of *generalized symmetry*. This principle states that all actants in an actor-network should be treated and can be described in the same terms. All differences between them are an *effect* of the relations in the actor-network and not given a priori. This bold statement of equating human and non-human actants has stirred a lot of debate. There are some serious problems concerning accountability and responsibility in heterogeneous networks, if you drop any difference between the human and the non-human, as I argue below. However, to my mind much of the discussion around the symmetry principle has been triggered by some misunderstandings about what is meant by *agency* and whether agency implies *intentionality*. While it might be possible and analytically fruitful to ascribe agency to the scallop, only few people would probably want to ascribe intentionality to the scallop in the process of its domestication. However one does not need to ascribe intentionality to the scallop in order to consider it to be an actant in epistemic processes. And without introducing the notion of intentionality when talking about agency, the ascription of agency might be more easily digestible for some of its critics.[64]

There is however, a more profound problem related to the principle of general symmetry and it concerns issues of *accountability and responsibility in heterogeneous networks*. These issues become particularly worrisome when applied to information systems. The main question here is the following: If human and non-human actants both are considered active agents in an actor-network and no difference between the two is assumed, who is to blame if things go wrong? Put differently if the "Waffen-Bürger" oder the "Bürger-Waffe", Latour's citizen-gun-hybrid is responsible for killing someone, to what extent can the citizen himself be made accountable for his action (Latour 2000: 218)? Is not the gun just as responsible? And what are the consequences

[64] For a sympathetic, but critical view on ANT, that portrays the major criticisms and misunderstandings in the reception of ANT in epistemology and philosophy of science please confer Longino (2002c).

of this distributed agency concerning the attribution of accountability? Does the rejection of any dualism between human and non-human actors imply that no-one can be held responsible for his or her actions anymore? Such blurring of analytical distinction concerning agency, as useful as it may be to counter trivializations or simplifications can have serious side effects. Leaving ANT-idiom for a moment, I would argue that since the attribution of accountability and the localization of responsibility is a major issue in our networked society. We have to be very explicit whether, when and to what purpose, we want to blur or abolish analytic categories and what the price is we are willing to pay for this. If the price is losing the possibility to take a critical stance towards existing socio-technical networks or not being able to localize responsibility anymore in case things go wrong, then I would argue that we pay too much. It might be very well, that ANT itself does not strive for such normative or critical goals anyway. But then each and everyone doing research has to decide how important these issues are for herself and where one has to depart from major claims of ANT for better reasons.

This issue relates to a more general point of critique concerning ANT's lack of normativity, i.e. its unwillingness to take any normative stance. Remaining at a purely descriptive level of analyses has been strongly criticized primarily by feminist scholars. They argue that when analyzing socio-technical systems with respect to inherent power relations, a normative standpoint has to be taken. This critique in combination with the uptake of many other claims of ANT, such as its anti-essentialist tendencies, has lead to a variety of feminist extensions and extrapolations of ANT within STS (cf. Bath 2009, Ilyes 2006) as well as to receptions within other fields of research (Longino 2002c).

Despite these points of critique, ANT has provided numerous important insights which are relevant for the questions of this thesis. First of all, by dropping the distinction between the social and the technical, the human and the non-human, ANT has helped exposing existing essentialisms with all their - possibly unintended - consequences. Further, ANT clearly has succeeded in the first anthropological dictum of "making the familiar strange", be it in their initial laboratory studies or in their deconstruction of familiar terms such as "society". Moreover by de-centering human agents ANT has put technologies, non-human living agents and materiality more generally into the spotlight. Even if this de-centering of humans may cause problems for locating responsibility if taken to an extreme, it has nonetheless been an instructive move that shed some light on actants that have been neglected by previous STS approaches.[65]

Another important insight of ANT concerns the *delegation* of social control to technology, as was famously demonstrated by Latour's accounts of the "Berliner Schluessel" (Latour 1996) and the door closer (Latour 1992). By showing how social control was substituted by technological artifacts, Latour and his colleagues have opened up a whole new way of looking at issues of power and their continuation in socio-technical networks. This issue of delegating control to technologies becomes relevant in Part 3 of this thesis, for instance when discussing the idea of "algorithmic authority" in the Web (Shirky 2009).

A further instructive aspect of ANT concerns its focus on the *processual* character of the actor-network, the *performativity* of socio-technical ensembles. Especially in relation to the major social epistemologies, which is depicted in Part 2, this acknowledgement of constant transformation and provisionality is illuminating and helps avoiding static conceptions of knowledge. Indeed Helen Longino, one of the social epistemologists who has been most sympathetic of certain insights of ANT, has also stressed the constant provisionality, partiality and plurality of knowledge (Longino 2002c). In the last years ANT has been used as an analytical framework to analyze many aspects and instances of information systems (e.g. Langlois 2005, Mager 2009 Tatnall 2003, Walsham 1997). One topic of particular importance has concerned the politics of search engines (c.f. Röhle 2009, as well as the seminal paper by Introna and Nissenbaum 2000). This fertility for the analysis of information systems should be another reason to take insights from ANT serious for analyzing epistemic social software.

3.1.4 Feminist STS

In the previous sections I have already indicated some of the major concerns various feminist have raised about central claims of ANT and the other STS approaches portrayed before. Thus, in the following, I focus on some of the conclusion they have drawn from this critique and sketch some of the decidedly feminist approaches to STS. Feminist concerns remain relevant throughout this thesis. However, in this section I

[65] For a cogent defense of taking materiality more serious, also confer Karen Barad's "Meeting the universe halfway" (Barad 2007), which is depicted below.

concentrate only on feminist responses to STS specifically and not on feminist epistemologies or feminist philosophical theories more generally.[66]

According to Nancy van House "[f]eminist STS is concerned with the daily, embodied practices of knowledge construction within historically changing structures and with power relations" (Van House 2003: 30). This quote renders the relevance of feminist STS for this thesis obvious. However, since the focus on practices, the acknowledgement of historical change as well as the focus on materialized and embodied forms of knowledge are also depicted in other approaches within STS, it is especially the critical analyses of the relationship between power and technology which strikes me as a crucial merit of feminist STS. Despite ANT's intention of analyzing power within actor-networks, ANT has been accused of being too uncritical towards unjust social structures by many feminist scholars. Thus, in contrast to other STS approaches, it has been mostly the feminist scholars who have problematized the relationship between *power and technology*. While early approaches often followed the simplified inscription-model of social structures being imparted onto technology, later approaches have also assumed more mutual relations between technologies and social structures (Van House 2003).

As becomes obvious from this focus on injustices, most feminist approaches are characterized by a distinctively *normative* – or even *political stance*. Instead of simply analyzing the ways in which social injustices are being inscribed into technology and reinforced or maybe rather undermined by technology, feminist scholars have been very creative and prolific in developing more constructive methods of how to design and develop socio-technical systems in more responsible and power-sensitive ways. This normative orientation is crucial for my own analyses and considerations on appropriate frameworks for analyzing and designing epistemic social software.

In the following I briefly introduce some aspects of the works of three feminist STS scholars who have been particularly influential: Donna Haraway, Karen Barad and Lucy Suchman.[67] I do not aim at giving any comprehensive overview over their works. Instead I highlight some of their crucial claims in so far as they are relevant for the

[66] It should be noted however, that especially Haraway and Barad deliver more general epistemological and ethical frameworks of analysis within their theories.
[67] I would like to thank Corinna Bath for bringing the works of Karen Barad to my attention as well as for sharing her knowledge on feminist STS and ICT in general.

topics of this thesis and have been neglected by other STS approaches. Beyond the relationships between power and technology these include issues of responsibility and accountability in entangled systems as well as performative understandings of socio-technological systems.

Donna Haraway: Networks, Dichotomies & Situated Knowledges

Donna Haraway has been among the first feminist scholars who have made ANT fruitful for gender studies. Criticizing ANT for its neglect of social inequalities and injustices rooted in existing power structures and being perpetuated by or even reinforced through socio-technical system design, she has developed her own network model. This model focuses on who and what gets excluded, by what means and for whose benefits. Haraway uses the reference to the string game "Cat's Cradle" to explicate her network model of knowledge and theory construction as a collective endeveaour, which is both "local and global, distributed and knotted together" (Haraway 1994, 70), hoping that it will help feminist scholars to "learn something about how worlds get made and unmade, and for whom" (Haraway 1994, 70).

According to Nancy van House (Van House 2003), there are two central issues that Donna Haraway has raised, which have been instrumental for STS and for the social studies of information and communication technologies in particular. One of them concerns her notion of *situated knowledges*, the other relates to her deconstruction of several classical dualisms. Such dualisms include mind and body, animal and machine, idealism and materialism, organism and machine, public and private, nature and culture, primitive and civilized (Haraway 1991b) and they all are easily relatable to another major dualism: female-male. Of central interest for this thesis is again the dualism of machine and organism, of humans and technologies. Many have seen the figure of the cyborg to be the central metaphor for the deconstruction of boundaries between humans and non-humans. Van House argues that this deconstruction of boundaries between us as human agents and the technologies and tools we use for knowledge construction and related epistemic practices can deliver new insights for information studies (Van House 2003). Concerning the goals and topics of this thesis, I would argue that dissolving a dichotomous understanding of *us as human epistemic agents* and *them as technological tools* is necessary to analyze and understand knowledge productive practices not only in science, but also in the socio-technical epistemic systems usually referred to as Web2.0

Haraway's concept of *situated knowledges* has clearly been among the most influential concepts in feminist epistemology. In her seminal paper "Situated knowledges: the science question in feminism and the privilege of partial perspective" (Haraway 1996). Haraway uses the metaphor of *vision* to delineate a feminist version of objectivity, that is able to remedy the shortcoming of previous feminist replies to claims of objectivity (Haraway 1996). According to Haraway most feminists including herself have either successively or even simultaneously attached themselves to two poles of feminist critique on prevalent accounts of objectivity: *radical constructivism* or *feminist critical empiricism*. Trying to avoid the pitfalls of both, in Haraway's definition, "[...] objectivity turns out to be about particular and specific embodiment and definitely not about the false vision promising transcendence of all limits and responsibility [...] only partial perspective promises objective vision" (Haraway 1996): 254). The concept of situated knowledges is thus indicative of the locality, plurality, particularity and embodiment (technical just as much as organic) of knowledge. A plurality, that technologically already get contested by my proof-reading software, whose red wiggly lines constantly reminds me that knowledge should better be used in its singular form.

The previous quote also introduces the notion of responsibility which is crucial in Donna Haraway's approach for two reasons. First, Haraway argues that proclaimed transcendence can be regarded as an effective way of shirking responsibility for one's epistemic practices. By contrast, acknowledging and disclosing locality enables the attribution of responsibility and accountability, a task that becomes all the more important, the more networked and computerized our society gets (Nissenbaum 1997). Haraway emphasizes this relationship between locality, responsibility and accountability in her essay on situated knowledges, which is "[...] an argument for situated and embodied knowledges and an argument against various forms of unlocatable, and so irresponsible, knowledge claims. Irresponsible means unable to be called into account." (Haraway 1996: 255).

However, there is a second way in which the notion of responsibility is crucial for Donna Haraway. Responsibility is not only something we want to ascribe to others, it is not only about making others responsible or blaming them. Rather Haraway insists on feminists' duty of taking "[...] responsibility for the social relations of science and technology" (Haraway 1991b: 181) Instead of a simple rejection or even demonization of technology, we have to take stakes at and participate in creating socio-technical worlds to live in. Remaining in a detached, analytic position of just revealing how social inequalities get inscribed into the techno-scientific is not sufficient. Rather, in her effort to take responsibility for the world we live in and to make them more livable, Haraway argues for an active political stance towards societal, resp. socio-technical change.

It is in this normative orientation that Haraway departs most clearly from ANT. This decidedly political stance, which does not retain itself to pure analysis and description is exemplified by her assertion that as interesting as it might be to analyze the human-non-human hybrids, her concern lies rather in the question "for whom and how these hybrids work" (Haraway 1997: 280, quoted after Bath 2009: 57). And this assertion can to a certain extent be read as a direct critique of Latour's work.

Moreover, by analyzing knowledge technologies with respect to issues of power, inequality and injustice, Haraway, as many other feminist epistemologists, relates her epistemology to ethics as well as social and political theory. And it is also one of the conclusions of this thesis that epistemology has to meet ethics and political theory when it comes to analyzing and amending socio-epistemic practices on the web.

Another normative consequence of Haraway's concept of situated knowledges concerns the necessity to include multiple voices into science and technology development. Since each and every perspective, each view is only partial, collaboration and the inclusion of many different, partial perspectives is an imperative, because it can provide a richer picture. However, this collaboration should not take the form of fusion nor does it imply that one could simply change one's perspective by lifting or dropping Rawls' *veil of ignorance* (Rawls 1971). It is rather the ideal of "[...] solidarity in politics and shared conversations in epistemology", that Haraway considers to be the alternative to relativism and totalization that a feminist account of objectivity could offer.

Despite her high appraisal, Donna Haraway has also received substantial criticism even within feminist theory. First of all, the Luddites amongst feminist theoreticians consider Haraway's stance towards technology not to be critical enough. The refusal of a stronger rejection of technology has led to frequent accusations by some feminist theoreticians that Haraway embraces technological developments too easily. However, this critique was probably a misinterpretation of Haraway, since she asserts that her writings are more *concerned* than optimistic and notes that it might be instructive to distrust our fears just as much as our hopes (Haraway 1996: 368f).

Another major point of critique that was raised against Donna Haraway is more crucial for the goals of this thesis however. It concerns the accusation that Haraway stays too abstract and narrative, that she refrains from making any concrete normative proposals about how exactly the technoscientific should be changed, how such more livable lives might look like and by which means they could be achieved. Her argument seems to be that any *recommendation* for system design would necessarily lead to *exclusions*. And since exclusions are what she trying to criticize and highlight, she remains rather silent on the practical implications of her analyses. Put differently, one of the major points of critique directed at Donna Haraway is that she is too much focused on text and too little on concrete action. As Wajcman phrases it: „At times, Haraway loses sense of how feminists could act to change, or at least redirect technologies, rather than reconfiguring them in our writings“ (Wajcman 2004: 101).

Karen Barad: Performativity, Matter & Responsibility

Another feminist theoretician, who has offered an interesting and original approach that shares certain core assumptions of STS, while reworking others is Karen Barad. Reading their frequent reference to each other, it becomes obvious that Haraway and Barad have been mutually influencing each other and share certain core assumptions. Such overlap concerns specific concepts, such as a posthumanism as opposed to anthropocentrism; shared focal points, such as hybridity and performativity; as well as certain goals, the most important of which are to take a critical stance and to take responsibility for the worlds we live in. As was case for Donna Haraway, Karen Barad's contribution to STS goes beyond an empirical assessment of the technoscientific and offers a new theoretical reflection, or rather a *diffraction* to use Barad's preferred term, on the processes involved in *knowing, being and, acting* (Barad 2007: 71ff).

More specifically, Barad's approach, which she labels *agential realism* (Barad 1998, Barad 2007), delivers an "[...] epistemological-ontological-ethical framework that provides an understanding of the role of human *and* nonhuman, material *and* discursive, and natural *and* cultural factors in scientific and other social-material practices" (Barad 2007: 26).

Agential realism offers a new way of understanding the relationship between the technical and the social, the discursive and the material, based on quantum physic, resp. a critical reception of Niels Bohr's philosophy-physics. Criticizing many popular truncations and misreadings of quantum mechanics, she argues for a more nuanced analysis of its possible philosophical implications (Barad 2007). According to Barad, one crucial point of Bohr's departure from classical mechanics concerns the unmaking of the Cartesian dualism of object and subject. In the process of physical measurement, the object and the observer, Barad's "agencies of observation", get constituted by and within the process itself and are not pre-defined entities. The results of measurements are thus neither fully constituted by any reality that is independent of its observation, nor by the methods or agents of observation alone. Rather, all of them, the observed, the observer and the practices, methods and instruments of observation are entangled in the process of what we call *reality*. For Barad, reality itself is nothing pre-defined, but something that develops and changes through epistemic practices, through the *interactions* of objects and agents of observation in the process of observation and measurement. Reality in this sense is a verb and not a noun.

However, *interaction* might be the wrong word to describe this entangled process, because *inter*action presupposes two separate entities to interact. Thus, to avoid this presupposed dualism, she introduces the neologism of "intra-action", to denote the processes taking place within the *object-observer-compound*, the entanglement of object and observer in the process of observation. This terminological innovation is meant to discursively challenge the prevalent dualisms of subject-object, nature-culture, human-technology and aims at opening up alternative, non-dichotomous understandings of technoscientific practices. Moreover, one of her major claims is that all those binary assumptions concerning the human *versus the* nonhuman, the material *versus the* discursive, and natural *versus the* cultural are fostered and hardened by prevalent, but misguided understandings of what matter, discourse, agency, objectivity, etc. mean. Thus, in developing her agential realism and to understand the relationship between matter and meaning, the material-discursive practices in science, she feels the need to rework many of these fundamental concepts - a challenge which she accepts in her most recent book (Barad 2007).

What should have become obvious from the previous portrayal is that Barad advances a *performative* view on the relationship between language and world as opposed to classical representationalism. Referring to Rouse (Rouse 1996), she argues that both, scientific realism and social constructivism share some basic assumptions concerning the relationship between reality and discourse and that those should be overcome by such a performative alternative, which focuses on practices and actions as opposed to correspondence (Barad 2003)). Barad's own *posthumanist performativity* is based on a critical reception of Judith Butler's concept of performativity (Butler 1993), which itself is inspired by Derrida reinterpretation of Austin's original differentiation between performatives and constatives. For Butler and others, performativity relates to the *ability of discourse to produce materiality*. She argues that social norms as well as categories such as sex or gender are pre-existent, but rather that they are continuously produced through processes of repetition and recitation (Butler 1993).

There is however, a crucial difference between Barad's and Butler's accounts. Barad accuses Butler of failing to show how "matter matters" and of granting language too much power (Barad 2003: 801ff). Moreover, she asserts that by concentrating too exclusively on the social or discursive, Butler is falling back into the Cartesian dualism of subject and object, in which matter is condemned to passivity. Thus, Barad's own performative account can be considered to be a materialist, naturalist and post-humanist transformation of Butler's original concept.

Taking matter serious and describing it as active, means to allow for non-human or hybrid forms of agency, a step that has been taken already with ANT's principle of general symmetry. However, in "Meeting the Universe Halfway", Barad sets off a section on agency and causality, with two quite illuminating quotes. At first Barad cites Haraway's stance towards the agency of the world by asserting that "[the world neither speaks itself nor disappears in favour of a master decoder. [...] Acknowledging the agency of the world in knowledge makes room for some unsettling possibilities, including a sense of the world's independent sense of humour. " (Haraway 1991a, cited from Barad 2003: 212). However, this quote is directly followed by a quote by Monica Casper, which criticizes that "[n]onhuman agency deflects attention from human accountability to other entities, whether human, nonhuman, cyborg, or what/whomever."(Casper 1994, cited from (Barad 2007: 213).

So here is the problem: If we attribute agency to non-human entities, can and should they be held responsible and accountable? Isn't that an invitation, a *carte blanche* to shirk responsibility? Do we let ourselves off the hook to easily and throw away any hopes for responsible and accountable actions?

My impression is that Barad's view on non-human agency and her stance towards ontological asymmetry has changed from earlier articulations (Barad 1996) to later ones (Barad 2007). In 1996, she still underscores the human role in representing, by stating that „[n]ature has agency, but it does not speak itself to the patient, unobstrusive observer listening for its cries – there is an important asymmetry with respect to agency: we do the representing and yet nature is not a passive blank slate awaiting our inscriptions, and to privilege the material or discursive is to forget the inseparability that characterizes phenomena" (Barad 1996: 181).

However, it seems that this special treatment of humans and especially the notion of *representing* does not well match her *posthumanist performativity*, as depicted some years later (Barad 2003). Finally, in "Meeting the Universe Halfway" ontological asymmetry is not even in the index and her answer to the question of non-human agency does not correspond to those earlier claims, but is a more nuanced dissolution of the distinction between human and non-human agency. By stating that "[a]gency is a matter of intra-acting; it is an enactment, not something that someone or something has" (Barad 2007: 216), Barad moves the locus of agency from singular entities to entangled material-discursive apparatuses. But even if agency is not tied to individual entities, it is bound with responsibility and accountability, as Barad makes very explicit in the following quote that can also be read as a response to Monica Casper's quote from before. "Learning how to intra-act responsibly within and as part of the world means understanding that we are not the only active beings— though this is never justification for deflecting that responsibility onto other entities. The acknowledgment of "nonhuman agency" does not lessen human accountability; on the contrary, it means that accountability requires that much more attentiveness to existing power asymmetries"(Barad 2007: 218f).

Thus, acknowledging that agency is not a human characteristic, but an attribution to certain phenomena within entangled networks *could* be seen as a *carte blanche* to shirk of responsibility. But this is clearly not the case for Barad. When developing her posthumanist ethics, Barad concludes that even if we are not the only ones who are or can be held responsible, our responsibility even greater than it would be if it were ours alone. She states "We (but not only "we humans") are always already responsible to the others with whom or which we are entangled, not through conscious intent but through the various ontological entanglements that materiality entails. What is on the other side of the agential cut is not separate from us—agential separability is not individuation. Ethics is therefore not about right response to a radically exterio/ized (sic!) other, but about responsibility and accountability for the lively relationalities of becoming of which we are a part."(Barad 2007: 393).

This focus on responsibility and accountability relates back to Barad's initial framing of agential realism as an "epistemological-ontological-ethical framework", a term by which she stresses the "[...] fundamental inseparability of epistemological, ontological, and ethical considerations" (Barad 2007: 26). Barad insists that we are responsible for what we know, and – as a consequence of her onto-epistemology for what *is* (Barad 2003: 829). Accountability and responsibility must be thought of in terms of what matters and what is excluded from mattering, what is known and what is not, what is and what is not.

This acknowledgement that knowledge always implies responsibility, not only renders issues of ethics and politics of such knowledge- and reality-creating processes indispensable. It also relates directly back to Barad's performativity. Epistemic practices are productive and different practices produce different phenomena. If our practices of knowing do not merely *represent* what is there, but shape and create what is and what will be there, talking about the extent to which knowledge is power or entails responsibility gets a whole different flavor.

For the purpose of this thesis, there are two aspects of Barad's work, which are highly relevant: Barad's notion of performativity on the one hand and her conclusions concerning agency, responsibility and accountability on the other. And these two aspects are indeed related. As Barad's states, "[a]gential realism is not about representations of an independent reality but about the real consequences, interventions, creative possibilities, and responsibilities of intra-acting within the world."(Barad 1998: 8). This concept of responsibility does not only enable the attribution of responsibility and accountability, it also is a request to take active part in science, in technology design and development, in those entangles processes that are characteristic of technoscience. It is our duty to watch out for power imbalances, injustices and the processes of silencing and exclusion in the agential cuts we make and we observe others making. We are to be held responsible and accountable for the consequences of our own practices and for those of others – human or non-human; we can be held responsible where we could have interfered, but did not. We may be held responsible and accountable for our practices and neglects even if we cannot guarantee or foresee how exactly they will affect the material-discursive apparatuses they are entangled in.

Finally, with respect to the analyses of epistemic social software, I would argue that Barad's posthumanist performativity might indeed be a useful tool for understanding not only how gender or reality materialize, but also of how different epistemic and social norms, ideas and practices get reinforced and re-enacted through practices.[68] Using a performative understanding of mattering also sheds a different light onto the questions of whether and why direct *inscriptions of values into technology* are impossible and why we do not have control over the acceptance, further development or the possible appropriations or even guerilla usages of the technologies developed. Any value-sensitive or explicitly feminist intention alone does not guarantee that structures of inequality are not still sustained or perpetuated in socio-technical systems.

Lucy Suchman: Expanding Frames & Accountable Cuts

The third feminist researcher in the broader field of STS, who has been particularly influential for analyses of information and communication technologies, is Lucy Suchman. As was the case for Donna Haraway and Karen Barad, there is considerable mutual influence between these three authors and despite certain differences, they share certain values, views and goals, such as a performative understanding of the socio-technical, the necessity for critical intervention and the relatedness of agency, responsibility and accountability. Suchman (Suchman 2007/2009) lists the major themes of her work in the preface to "Human-Machine-Reconfigurations", a reprint of her seminal book "Plans and Situated Actions" (Suchman 1987), amended by new footnotes and some additional chapters that account for theoretical and empirical developments that have taken place in the 20 years in between these two publications. These major themes are: "the irreducibility of lived practice, embodied and enacted; the value of empirical investigation over categorical debate; the displacement of reason from a position of supremacy to one among many ways of knowing in acting; the heterogeneous socio-materiality and real-time contingency of performance; and the new

[68] The emphasis on the performative aspect of epistemic practices is also crucial for Martin Kusch's notion of performative testimony (Kusch 2002), which is be portrayed in Chapter 5.

agencies and accountabilities effected through reconfigured relations of human and machine" (Suchman 2007/2009: xii).

Suchman's work is of particular interest for this thesis, because she applies feminist STS and also explicitly Karen Barad's agential realism to the design and development of information technology. Suchman is now professor of anthropology of science and technology, but before turning to academia, she has worked as a researcher for 20 years at XEROX PARC. As a trained anthropologist, whose approach is characterized by ethnomethodology and conversation analyses, Suchman is a decided empiricists in the sense that she values empirical analyses over abstract conceptual debates (Suchman 2007/2009: xii). Since this thesis is rather conceptual, it seems to be at odds with Suchman's own focus on empirical analyses. However, I take Suchman's focus on empirical analyses not to be an argument against conceptual work *per se* but rather as a critique of those conceptual approaches that simply ignore empirical evidence, a critique that I share with her. Moreover, since my goal is that some of the conceptual insights from this thesis will find their way back into practice, i.e. into technology design and development, I take the methodological conclusions that she draws from her own analyses to be crucial. I return to this at the end of the section. And finally, there also is a clear link between the empirical and the conceptual in Lucy Suchman's own work. At PARC, she empirically investigated human-computer interaction (HCI) based on ethnomethodology and conversation analysis. But the results she obtained with her anthropological, empirical research made her question and criticize some of the key concepts in the field of HCI itself, namely the notion of the *interface* itself, the role of *plans* and the *cognitive foundations* of HCI. Thus, in her seminal "Plans and Situated Action", Suchman re-conceptualized the conceptual basis of HCI through her empirical analyses (Suchman 1987).

To being with, Suchman questions the common understanding one of the most important concepts in HCI, namely the *interface*. Instead of regarding the interface as a fixed and stable border between humans and machines, through which messages have to be sent, Suchman takes "[...] the boundaries between persons and machines to be discursively and materially enacted rather than naturally effected" (Suchman 2007/2009: 12). This means that first of all, she is also proposing a *performative view on the socio-technical*. In referring back to Judith Butler's argument that sexed and gendered bodies are materializations of repeated and recited gender norms, she argues that “[t]echnologies, like bodies, are both produced and destabilized in the course of these reiterations” (Suchman 2007/2009: 272).

Moreover, such a view on the interface as being *enacted* also challenges any radical boundary setting between humans and machines. Instead of dwelling on the proposed differences or similarities between humans and machines, the question for Suchman has turned out to be rather “[...] how and when the categories of human or machine become relevant, how relations of sameness or difference between them are enacted on particular occasions, and with what discursive and material consequences” (Suchman 2007/2009: 2). Thus instead of sharply separating the human and the machine and discussing issues of agency and responsibility within this dichotomous framework, Suchman joins Barad and Haraway in understanding the socio-material as *entangled* (Barad 2007), as a *material-semiotic compound with distributed agency*, that gets re-enacted through repetition and re-citation.

What is crucial about this emphasis on the performative, dynamic and entangled nature of the sociotechnical is that Suchman does not equate humans and machines. Quite on the contrary, she criticizes that the whole language of interactivity in HCI conceals the crucial differences between humans and machines, differences which people experience in their encounters with machines. Her emphasis on the asymmetry between humans and computer however, is not based on some presupposed ideology of human supremacy. Rather, this difference is a conclusion drawn from her empirical investigation of the *interaction patterns between users and technological artifacts*. The empirical basis of her book “Plans and Situated Action” (Suchman 1987) were analyses of the interaction between employees at PARC with a new multi-function photocopier. In the analysis of such videotaped encounters between the photocopier and the PARC employees, who were probably not the most technically illiterate test group, Suchman made out crucial differences between human-computer interaction and human-human interaction.

One of the key insights that she transferred from anthropological conversation analyses to HCI was that communication and interaction among humans does not follow a simple sender-receiver model of information transfer. However, during the early 1980ies, such a model was a common framework for many conceptualizations in HCI as well as in Artificial intelligence (AI). Opposing such a simplified model of communication, she argued that “humans dynamically construct the mutual intelligibility of a conversation through an extraordinarily rich array of embodied interactional competencies, strongly situated in the circumstances at hand [...]" (Suchman 2007/2009: 10). If one adopts such a model of communication, it becomes obvious that machines by contrast, have access to much more limited information about humans as their interaction partners. This observation clearly becomes vivid if one only thinks about the reduction of available information that already occurs when *human-to-human* interactions are conducted via email or SMS and a lot of information, including all the non-verbal signals gets lost. Thus, Suchman argues, there is no such thing as *interaction* between humans and machines achieved yet, especially not if one considers interaction in a stronger sense to be "[...] a name for the ongoing, *contingent coproduction* of a shared sociomaterial world" (Suchman 2007/2009: 23, italics in original). She concludes that, interface design has to adopt different strategies to accommodate for these insights. The crucial twist she proposed was to reframe "[...] the problem from creating a self-evident machine (or one able to engage in interaction with its user), to writing a user interface that is readable, with all the problematics that reading and writing imply" (Suchman, Blomberg et al. 1999: 395, quoted from Van House 2003: 35).

The second crucial theoretical shift of Suchman's analyses is indicated by the title of her book "Plans and Situated Action" (Suchman 1987). It concerns the reconceptualization of the role of plans in reasoning and action, which entails a critique of the standard model of human reasoning and action. By considering plans to be *tools and parts of human action* rather than "cognitive control structures that universally precede and determine actions" (Suchman 2007/2009: 13), Suchman has stressed the embodied and enacted character of knowing in action. Despite frequent misinterpretations, Suchman clarifies that she never argued that plans are not important for action or that they are in oppositions to (contingent) action. Rather she considers plans to be a special kind of action and is interested in "[...] the relation between the activity of planning and the conduct of actions-according-to-plan" (Suchman 2007/2009: 21).

This shift in understanding a) rational action not to be the result of a plan developed and then carried out by a secluded reasoner, and b) considering this reasoner not as a "bounded, rational entity but [rather] as an unfolding, shifting biography of culturally specific experience and relations, inflected for each of us in uniquely particular ways" (Suchman 2007/2009: 23) challenged some of the cognitive, or cognitivistic foundations of HCI as well as AI at that time. Such a model of intelligent action is in opposition to a classical model of a detached reasoner who has escaped the contingencies of time, space, technology, materiality or discourse, the "the heterogeneous sociomateriality and real-time contingency of performance", as she calls it (Suchman 2007/2009: xii).

Given the socio-epistemological framework of this thesis, this proposal is a particular relevance, because this picture of a reasoner as an autonomous, rational being that ponders in solitude, outside of time and space was not only characteristic of AI and HIC, but has also been the standard model of a reasoner in philosophy. Although, social epistemology aims at correcting this picture many authors even within social epistemology are still in the grip of this individualistic view, as is shown in the next chapters. Thus, a crucial insight to be obtained from Lucy Suchmans's work concerns this departure from a monadic view of reasoning and the embrace of a more situated, relational view on cognition and reasoning.

As noted before *Human-Machine-Reconfigurations* (Suchman 2007/2009) amends the original publications of *Plans and Situated Actions* (Suchman 1987) by some additional chapters. In these chapters Suchman takes into account the work that has been done in the twenty years between those two publications. For Suchman, taking these insights from STS, especially feminist STS, but also from other fields within the social science and humanities serious has two important consequences for studying the socio-technical. And these methodological consequences culminate in two research imperatives: *expand frames and make accountable cuts* (Suchman 2007/2009)!

What does Suchman mean by cutting and framing? Suchman's usage of the term *cut* refers back to Barad *agential cuts* introduced in the previous section. Focusing on cuts, Suchman points our attention to *how objects of analysis are created by being cut out of entangled networks*. Consistent with her critique of individualistic ontologies, she argues that objects of analyses or intervention are not *given*, but *made* by different cuts into the socio-material assemblages. Such boundary work, the making of differences, is a basic foundation of reasoning. There is no sense-making without making differences, without separating some and joining others, an insight well documented in cognitive science and psychology. However, as necessary and unavoidable, as implicit and unnoticed this process of differentiating and cutting is – it is always also a political act. Every cut, every distinction includes some and excludes others – and every inclusion and exclusion is consequential for the socio-material assemblage as a whole. Referring back to Karen

Barad's agential realism, Suchman notes that "because boundaries have real consequences, accountability is mandatory" Suchman 2007/2009: 285, quoting Barad 2007: 187).

However, acknowledging the relational and entangled nature of the sociomaterial, implies that agency cannot be localized in individual entities, but rather is distributed within socio-material assemblages. As Suchman notes, again resonating Barad "[...] agencies – and associated accountabilities – reside neither in us nor in our artifacts but in our intra-actions" (Suchman 2007/2009: 285). The question, however, remains how exactly to be responsible, how to hold or to be held accountable if agency is distributed. How can we maintain responsibility and accountability in such a networked, dynamic and relational matrix? Although I think that Suchman goes into the right direction, she remains quite vague about this in her concluding remarks of *Human-Machine-Reconfigurations*. She states that "[r]esponsibility on this view is met neither through control nor abdication but in ongoing practical, critical, and generative acts of engagement. The point in the end is not to assign agency either to persons or to things but to identify the materialization of subjects, objects, and the relations between them as an effect, more and less durable and contestable, of ongoing sociomaterial practices" (Suchman 2007/2009: 286). And one of the crucial tasks in this enterprise consists in *extending the frames of analysis* "[...] to metaphorically zoom out to a wider view that at once acknowledges the magic of the effects created while explicating the hidden labors and unruly contingencies that exceed its bounds"(Suchman 2007/2009: 283f).

Suchman's work is instructive for this thesis for several reasons. First of all, Suchman's performative, dynamic and relational understanding of the socio-technical can help to understand the ways in which certain epistemic methods, values and norms are being reinforced through repeated socio-technical epistemic practices. Moreover, Suchman's treats socio-material assemblages as *objects of analysis and intervention*. That is she focuses on critical analyses just as much as on conclusions concerning the design and development of technologies. In line with her explicit request for critical technology design, she offers conclusions on how to accommodate for these insights for technology design and development. In particular this refers to her request to expand frames, to reflect and question frameworks of modeling, to develop new alternative models, to open up new perspectives and enable re-configurations of the socio-material.

Another crucial aspect concerns the difference between humans and machines. Suchman acknowledges that it has been important to tear down the proclaimed boundaries between humans and machines and that this has opened up new ways of thinking about their relationship in a more performative and dynamic way. However, the time for her seems to be ready to go a step further, to make another "agential cut" in order to get a grip on the *consequences* of these differences. Her notion of *accountable cuts* is a reminder that even if making cuts is unavoidable, boundaries have consequences and that accordingly we are responsible and accountable for the cuts we make. Crucially, for Suchman being aware of one's cuts and frames as well as the duty to expand frames and make accountable cuts is not only the task for some critical social scientists or feminist scholars. Rather it is the technology designers and developers themselves who have to expand their frames and be accountable for the cuts they make. Drawing lines, making boundaries are basic tasks in our daily acting in the world just as much as they are basic tasks for designers of technologies. But those cuts are never innocent, they have consequences, they include and exclude, and we are to be held accountable for the consequences of our actions even if we cannot foresee them.

Adopting such a dynamic and entangled view on socio-material assemblages may lead to the conclusion that since consequences of our interventions are not controllable, there is no need for critical intervention. After all, one of the lessons learned from Suchman, but also from other feminist STS approaches portrayed before is that if we take performativity and relationality into account, it is futile to assume that we could *directly inscribe* certain values to achieve certain effects. However, this insight does not mean that we can let ourselves off the hook too easily. On the one hand, the critical analytic tasks of analyzing effects of certain interventions, of expanding the frames, of opening black boxes and conceiving alternative models for technology design remains. Moreover, as Suchman reminds us, the solution for this dilemma of distributed agency in entangled socio-material assemblages can only lie in "[...] the processes ongoing practical, critical, and generative acts of engagement" (Suchman 2007/2009, 286) with the world we live in. Thus, there is room for critical analysis and critical technology development – and a request to join this endeavor.

3.2 Values and the Politics of Technology

In this chapter I have briefly sketched the major approaches in STS as well as some feminist critiques and extensions of STS that appeared particularly relevant for this thesis. To conclude this chapter I would like to return to one major issue that has been underlying many previously outlined debates. It is the relationship between the societal values and technologies. The guiding questions for this section are: What do we mean by saying that societal values are inscribed into technology? Or looking the other way: What do we mean by saying that technology might influence and change societal values?

To approach these issues, we first of all have to address the question of what values are. The second crucial question concerns the procedural character of this relationship between the technical and the societal. Taking claims about the performativity of the socio-technical serious we have to ask whether the notion of *inscription* might not be misleading in suggesting the existence and interaction of two disparate entities, thus fostering a dichotomy that should rather be abandoned. Besides such constitutive questions around the relationship between technologies and societal values, there are also questions concerning the *consequences* of any such conceptualization. For instance, can the inscription of values, the role values play in technology design be avoided? Should they be avoided? Is it even possible to promote a responsible and accountable design that takes the insights from STS serious? To open up this debate on the relationship between values and technologies I return to the beginning of STS, to one of its funding texts: Langdon Winner's "Do artifacts have politics?" (Winner 1980).

3.2.1 Inscription and The Politics of Artifacts

The Starting Point: Winner's Bridges

In this widely cited article, Winner argues that technologies are by no means neutral, but that they have political properties by embodying "[...] specific forms of power and authority" (Winner 1980 121). Referring back to Lewis Mumford's differentiation between authoritarian and democratic technologies, Winner offers a diversity of examples to support his claim that artifacts have politics. While the political nature of the atom bomb may be straightforward, Winner's other examples appear much more innocent at first sight: the mechanical tomato harvester, cotton-spinning mills, automobile assembly teams, and Baron Haussmann's re-structuring of Paris. Given the number of examples in a quite short article, it should not come as a surprise that all those examples are only briefly introduced and none of them reaches the level of a full case study – and neither does his most famous - or infamous – example: Robert Moses's parkway bridges in New York.

Winner's empirical starting point for his analyses on the politics of artifacts has been the observation that the parkway bridges in New York are "extraordinarily low" (Winner 1980: 123). The person in charge of building those bridges was Robert Moses, "[...] legendary political entrepreneur, who has shaped the physical form of New York in this century and beyond as no other person" (Joerges 1999: 412). Departing from this seemingly innocent empirical observation about the height of the parkway bridges, Winner argues that Moses *intentionally* had those bridges built that low to "discourage the presence of buses on his parkways". By this trick he was able to "[...] limit access of racial minorities and low-income groups to Jones Beach, Moses's widely acclaimed public park" (Winner 1980: 124). Winner argues that the design of those parkway bridges reflects "[...] Moses's social-class bias and racial prejudice" (Winner 1980 123) and concludes: "Many of his monumental structures of concrete and steel embody a systematic social inequality, a way of engineering relationships among people that, after a time, becomes just another part of the landscape."(Winner 1980: 124).

Winner argues that due to his social and political power as a city planner, Moses was able to *inscribe* his race and class related prejudices into the design of parkway bridges. By making them too low for buses to drive through, he was able to prevent black people as well as other low-income groups dependent on public transportation, from accessing these most scenic routes and the places where they led. To put it in a nutshell, Winner argues that Moses intentionally inscribed his values and prejudices into technology in order to achieve certain societal effects.

The Refutation Which Did Not Matter

Almost 20 years later, this seemingly straightforward story was refuted by Bernward Joerges, who claimed that Winner's story about the parkway bridges while being a "highly successful parable" (Joerges 1999: 416), unfortunately is counterfactual. Based on correspondences with US civil engineers Joerges asserts that commercial traffic was forbidden on the parkways in general and that since the transport situation on Long

Island was already good, there was no reason to waste the money on building higher bridges. Hence, Joerges concludes that "*Moses could hardly have let buses on his parkways, even if he had wanted differently*"(Joerges 1999: 419, italics in original).

Irrespective of these methodological flaws and the questionable conclusions, Winner's example is a success story. It's been recited in almost all account of STS, including this thesis and keeps inspiring research until nowadays. How is that possible? Despite his thorough critique and his analyses of Winner's story as a rhetorical device, Joerges himself already concludes that Winner's story serves a purpose: " to resituate positions in the old debate about the control of social processes via buildings and other technical artifacts – or more generally, about material form and social content"(Joerges 1999: 411).

To understand how Winner's story – for better or worse – has become so successful, one has to contextualize the story itself. Originally published in 1980, this simple, rhetorically brilliant, short and easily transferrable parable about technologies in society was very seductive for many reasons. As Joerges summarizes: "Enormous interpretive flexibility, unambiguous empirical reference, elegant theoretical formula. And all that coupled with an urgent political-moral message. Not bad." (Joerges 1999 420f).

What was so promising and inspiring about Winner's case is that he delivered a simple and strong case for the *inscription of societal values into technology and the societal effects of such biased technologies*. It is this insistence on the political character of artifacts and the possibility of social engineering through technology that hit the Zeitgeist of critical science and technology scholars. Winner initiated a discussion about the politics of artifacts by refuting the assumption that technologies are neutral or follow some inner-technological rationality. Instead he stressed the societal environment with all its values, prejudices and assumptions that get inscribed into these artifacts. In Moses case – and that makes this specific example even more seductive, there seemed to have been this powerful man who *intentionally* inscribed his views into technology, who quite literally carved his racial prejudices and societal inequalities into stone, made them durable and solidified them in artifacts, and ensured their enduring societal impact.

Hence, it is not the validity of the claims of the article itself that explain its continued relevance, but the resonance that it caused. This is the reason why especially feminist scholars in STS still relate back to Winner's article, despite the refutation of the particular case of Moses through Joerges. Winner's article has been and remains to be a starting point for considering artifacts to be potentially or actually political by showing *how certain values get inscribed into technology and how such biased technologies have societal consequences*. Almost all crucial issues of STS can be found in this short parable of the New York parkway bridges: social inequalities and injustices, inscription of societal values into technological artifacts, subsequent blackboxing and concrete societal effects of technologies. Thus, irrespective of its counterfactuality, Winner's article offered all ingredients for a perfect case study. And the notion of *inscription* in particular has been thoroughly influential.

Inscription: From Winner to ANT and Feminist STS

Despite many pronounced differences between Winner and ANT, the concept of inscription itself is a central concept in ANT as well (Akrich and Latour 1992). It has been employed in particular in Madeleine Akrich's analyses. Her *script model* has made the concept of inscription, as well as the processes of *de-scription, transcription* and *re-inscription* fruitful (Akrich 1992). Akrich argues that "[a] large part of the work of innovators is that of "*inscribing*" their vision of (or prediction about) the world in the technical content of a new object" (Akrich 1992: 208). She calls the end product of this work a *script* and argues that "[...] like a film script, technical objects define a framework of action together with the actors and the space in which they are supposed to act" (Akrich 1992: 208).

Both Akrich and Winner claim to avoid social determinism as much as technological determinism, but they follow different lines of reasoning for this. One crucial difference between Winner and Akrich, is that while Winner stresses the intentions and actions of the master designers, such as Robert Moses, Akrich emphasizes the interplay between designers' plans and the users' actions, the interaction between "[...] *the world inscribed in the object and the world described by its displacement* (Akrich 1992: 209). Thus, by focusing on the possibilities of adjustment and appropriation through the users, Akrich offers a more nuanced version of the respective relevance of designers and users.

One may read Akrich's case studies about the appropriation of technological artifacts, such as the photoelectric lighting, as arguments that the innovators visions and intentions should not be overrated, because the outcome depends on usage in the end. However, I think this interpretation goes too far, because Akrich herself notes that *it is* *only in usage contexts which are quite different from the developmental contexts that such appropriation occurs.* In the majority of cases, namely those situations in which the context of development and the context of usage are more similar however, the scripts of technical objects remain crucial, because they set the limits and affordances of the usage of the technical objects. This means that although usage is relevant and appropriation possible, design decisions remain important, or in Akrich's own words: "[...] although users add their own interpretations, so long as the circumstances in which the device is used do not diverge too radically from those predicted by the designers, it is likely that the script will become a major element for interpreting interaction between the object and its users" (Akrich 1992: 216).

After this inscription has taken place, the artifacts are closed and the processes of development are blackboxed. Reminiscent of Winner, Akrich concludes her analyses with the following remark concerning the politics of artifacts: "This is why it makes sense to say that technical objects have political strength. They may change social relations, but they also stabilize, naturalize, depoliticize, and translate these into other media. After the event, the processes involved in building up technical objects are concealed. The causal links they established are naturalized. There was, or so it seems, never any possibility that it could have been otherwise" (Akrich 1992: 222).

It should not come as a surprise that feminist scholars have adopted the idea of an *inscription* of societal values into technologies and their influence back onto society and societal values. More specifically, feminist scholars appropriated Winner's claims as well as Akrich's script model to analyze and show how gender stereotypes and related inequalities and injustices are inscribed into and reinforced through technology. One of the most influential extensions of Akrich's script model is the concept of *gender scripts*, with Ellen van Oost, Nelly Oudshoorn and Els Rommes as the main proponents (cf. Bath 2009). The gender script model allows analyzing how gender stereotypes get into technologies and in contrast to Akrich's original model does not only focus on the designers and the users of technologies. Rather it puts the broader societal environment into perspective and highlights other stakeholders as well as *non-users*. With the reference to non-users is becomes obvious that the gender script model focuses on inclusions and exclusions, a move that has been argued for in other feminist approaches portrayed before. Another crucial point about the gender script model, in which it explicitly departs from Winner concerns the intentionality of designers. As Els Rommes notes, “[...] a gender script will rarely be the result of conscious attempts of designers to exclude certain users. Rather, it will be the result of unconscious repetitions and reiterations of the hegemonic masculine norm” (Rommes 2002: 19, quoted from Bath 2009: 83).

What conclusions can be drawn from Winner, Akrich’s and their feminist extension concerning the inscription of values or the role values play for the socio-technical? What are the consequences for this thesis? First of all, we should acknowledge the political character of technological artifacts. Decisions made in technological design and development have consequences that go beyond the look-and-feel of technological artifacts. Such decisions set affordances and limits. Even if artifacts can be used in ways other than intended by the designers, if defaults can be changed or subverted, they deliver the keys for interpretation, which have to be actively circumvented to be undone. Further, we have seen that the intentionality of the designer is not a necessary prerequisite for the politics of technologies. Indeed, this may even be the exception and most inscriptions of values and stereotypes may occur rather implicit. Finally, the term inscription should be used with caution, since it implies not only intentionality, but also a directness of societal values implemented in technologies leading to societal effects that contradicts insights concerning the entangled nature of the socio-technical.

3.3 The Pragmatic Turn: Values in Design

If these are the analytical insights from Winner and STS more generally, what constructive conclusions can be drawn from them? If we consider technologies to be in principle and possibly ineluctably political, what are the consequences to be drawn from this observation? Do we have to remain on the analytical and critical side, disclosing in what ways and with which consequences, artifacts are biased? Are we confined to the role of those crying “wolf!” if some questionable values are being inscribed into technology, which might have detrimental effects for different groups users and non-users?

Despite my deep conviction, that such a watchdog role, will remain crucial for critical scholars of technologies, this is not the only imaginable role. One at least equally important task will be to develop guidelines and recommendations of how to develop *value-sensitive technologies,* i.e. technologies that not only function well, but that take the insights obtained from critical STS serious. Thus, instead of only revealing, which biases and prejudices have been inscribed into technologies and demonstrating their negative consequences different agents, it should be possible to play a more *constructive role within the process of technology design and development.* According to Flanagan, Howe et al. such a "[...] pragmatic turn [...] sets forth values as a design aspiration, exhorting designers and producers to include values, purposively, in the set of criteria by which the excellence of technologies is judged" (Flanagan, Howe et al. 2008: 322).

Accordingly, in the remaining part of this chapter, I focus on approaches that try to bridge the gap between critical analysis and reflection and the development of technological artifacts. To describe these approaches I use the label *Values in Design*. *Values in Design* as conceived here is not a clear-cut program with a distinct set of methods, theories or scholars. Its roots lie in STS, just as much as in applied ethics and critical design practices within computer science and the term is rather meant to refer to a broader set of approaches that twists the insights obtained from STS and critical technology studies into developing guidelines or recommendations for technology design.

The publication of the seminal book *"*Human Values and the Design of Computer Technology" edited by Batya Friedman can surely be seen as a catalyst for the pragmatic or constructive turn in debates around values in design and may thus serve as a vantage point for this portrayal (Friedman 1997a). For this anthology, which marks one of the major points of departure for any systematic analysis of values in design in the field of computer science itself, Friedman brought together an interdisciplinary group of acclaimed scholars tackling the issues around values in computer and information system design. In her introduction, she asserts that although designers hardly think about values in their daily business, they "[...] necessarily impart social and moral values" (Friedman 1997b: 1). But if that's the case, she further asks: "Yet how? What values? Whose values? For if human values – such as freedom of speech, rights to property, accountability, privacy, and autonomy- are controversial, then on what basis do some values override others in the design of, say, hardware, algorithms, and databases?" (Friedman 1997b: 1).

In her anthology, Friedman focuses on *moral values*, which she sets aside from *personal* and *conventional values*. She grounds this distinction between moral values and conventional values in certain psychological literature, where moral knowledge is distinguished from conventional and personal knowledge (e.g. Nucci 1996; Smetana 1983, both cited from Friedman 1997b). To my mind, this distinction is problematic taking into account that moral values can be considered to be societal conventions themselves. Moreover, even the briefest look at information systems design reveals that how values are effectively introduced and stabilized through standardization, a paradigmatic case of societal conventions. Indeed, within Friedman's anthology, the debate between Lucy Suchman, Terry Winograd, and Thomas W. Malone is a clear indicator of the import and "value-laddenness" of standardization, classification and the definition of categories (Suchman 1997, Malone 1997, Winograd 1997). More generally, it has been particularly the works of Geoffrey Bowker and Susan Leigh Star that have clearly shown the import classification and standardization and the values at play in them (Bowker and Star 1999). Hence before getting into the details of how to make use of values for information systems design, it may be necessary to take a closer look at what is meant by *values*.

3.3.1 Approaching Values

Throughout this whole chapter I have implicitly talked about values. Different STS scholars have emphasized the inscription of certain societal values into technologies, the ways in which such values get reinforced by technologies, how they are recitated and stabilized in socio-technical systems, the ways in which technologies might change societal values, etc.. Despite its manifold appearance, the term *value* did never get properly introduced in this chapter. Before the hopes get too high, let me interject, that I do not deliver any substantial let alone any comprehensive introduction to the topic of values. The reason for this – again – is that the topic is too broad to be just slipped into a subsection of this thesis. Even if only applied to the fields of science or technology, the discussions around the topic of values are broad, deep and diverse.

What are values? The Merriam-Webster Online Dictionary offers seven different definitions of value, ranging from a "fair return or equivalent in goods, services, or money for something exchanged", over the monetary or relative worth of something as well as the concepts of value in mathematics and music all the way to an understanding of value as "something (as a principle or quality) intrinsically valuable or desirable" (Merriam-Webster-Online-Dictionary 2009).

As should become obvious from the definition above, the concept of values is not only widely, but also diversely used in our society. Philosophers have discussed the intension as well as the extension of values, the differences between shared values and personal values, between values as norms and values as feelings to name just a few examples (Mitcham 2005). Despite such philosophical debates around values, there are also other disciplinary perspectives on values. In an article on values and valuing in the *Encyclopedia of Science, Technology, and Ethics*, Mitcham differentiates between economic, social scientific and philosophical perspectives on values (Mitcham 2005). Accordingly, although people use the notion of values frequently in their daily life, what is meant by values is far from clear and differs profoundly between and even within disciplines.

Clearly, not all of these issues can be dealt with in this thesis, so some focusing is necessary here. Given the topic of this thesis, I focus on definitions of values within the field of *Values in Design* itself, which also sets the frame for this section. For their analyses and methodological recommendations concerning value sensitive design of information systems, Friedman and her colleagues, for instance, define values quite broadly as something that "[...] a person or group of people consider important in life" (Friedman, Kahn et al. 2006: 349). They relate their definition to the Oxford English Dictionary definition of value as "[...] the principles or standards of a person or society, the personal or societal judgement of what is valuable and important in life" (Simpson and Weiner 1989, quoted after Friedman, Kahn et al. 2006: 349). In many articles on values and technologies, authors suspend from giving definitions of values altogether and confine themselves to giving some examples or to list those values they consider relevant for information systems.

For this thesis, the following specifications are made. First of all, while one might clearly argue that economic values also play a role for social software, it is rather the social science and philosophical perspectives that is be relevant for this thesis. More concisely, it is the *sociological concepts of shared societal values* as well as *the ethical examination of values* that are most relevant in this section. Moreover, despite the role that *personal* values of designer might play in the development of technologies, the focus of this thesis is on societal values, i.e. *values as socially shared judgments about what and who is how important.* This understanding of societal values as *shared* judgments does not imply that values are *universally shared* - quite on the contrary.

Values are always someone's values, the shared judgments of some group and not another. And indeed, the fact that values differ between groups explains why despite the fact that values are predominantly considered to be a positive notion, they have been the introduced as something to be *avoided* in science as well as in technology design and development.

If values are shared judgments about what is important in life, it is clear that such judgments might differ between different people. And then the question arises *whose values* are talked about, whose values are relevant and whose values get ignored or even antagonized. When values are inscribed – irrespective of good or bad intentions, it is always a process of imposing one's view on others. And in the case of technology, these values are not only made invisible, they are also solidified in the artifacts. Hence, it is this combination of invisibility and enduring impact, which makes a proper analysis of values in technologies all the more relevant.

With respect to science, Mitcham indicates that there is a bidirectional relationship between values and science, by stating that "[v]alues and valuing are as much a challenge to science as science is to values" (Mitcham 2005). In adopting the same view on technology, this would mean that not only do societal values influence technology design. Technologies themselves also influence and may change societal values. Thus, the moment we take part in technology design, we are to be held responsible and accountable for the decision we make, because they might have societal consequences. The crucial question therefore has to be how we can account for the critical insights obtained from science and technology studies while also accounting for and taking part in the nitty-gritty of technology design and development. In the following I give two examples of analyses of information and communication technologies with respect to the values of *accountability* and *freedom from bias*. I also portray two methodological approaches of how to consider values in the process of technology design developed by Batya Friedman as well as Helen Nissenbaum and her colleagues. But before turning to the examples I want to return the topic of intentionality and its relation to values.

3.3.2 Intentionality Revisited

In the previous sections, I have argued that intentionality is not necessary for values to be inscribed into technology. However, the opposite claim, that the intentional inscription of values is impossible does not hold true. Although intentionality is not *necessary* for values to be inscribed into technological artifacts, it clearly is *possible* to inscribe one's values intentionally into technologies. Thus, it is of course imaginable that some racist, sexist programmer develops a tool that discriminates systematically against women and people of different race. But if we grant that some evil designer might use his or her technological artifact for some evil social engineering, we must also allow for the possibility to inscribe desirable values into technologies.

However, one of the lessons learned from the previous sections of this chapter is that the *results* of any attempt of social engineering by explicitly inscribing certain values into technology are far from clear. Taking the role of the users, the possibilities of appropriation and subversion serious, we have to acknowledge that there is no direct link between technology design and societal effects. If this is true for the negative examples of biases and prejudices, of the inscription of societal inequalities and injustices into technologies, then the same must be true for any attempt to change the world for the better with the help of technologies. As critical scholars just as much as designers it is our duty to reflect upon the values we wish to inscribe into the technologies we develop. Unfortunately, there is no guarantee that even our best intention cannot be subverted. But, as insisted before, this recognition of the performative and entangled character of the socio-technical does not diminish the responsibility of designers – it just puts this accountability and responsibility into context. And the field that I consider to be most promising in that respect is *Values in Design*. Let's have a look how scholars in the field of *values and design* have tackled that task of being critical analysts and constructive designers at the same time.

3.3.3 Which Values? Two Examples for Socio-Technical Systems

In this section I want to give two examples of values that play a role in computer technology and information systems design. The two values that I focus on are *accountability* and *freedom from bias*, because I consider them to be among the most relevant ones for this thesis.

Accountability was chosen, because in the previous sections, most notably in those on the feminist approaches in STS, accountability and responsibility have been identified as crucial aspects for critical analyses of socio-technical systems. Moreover, it has been argued that in entangled and networked environments, finding the locus of accountability and responsibility is not only more important, but unfortunately also more difficult. Since the examples of epistemic social software depicted in the previous chapter are a clear case of such entangled networks with distributed agency, issues of accountability and responsibility are of particular concern.

Similar reasons made me chose *freedom from bias* as the second values I want to put stress on. Analyses conducted by feminist scholars in particular have emphasized that in many socio-technical systems, different groups of people are affected differently, that design decision have different consequences for people depending on the groups to which they belong. As has been noted before, it might not be possible to ensure that no-one is ever treated unfairly and even with the best intentions, discrimination against some people or others can occur. However, this does not diminish the duty to watch out for such injustices and to propose ideas of how to avoid them.

Accountability

In an early essay on values and computer technology, Helen Nissenbaum warns from the danger that in a society as pervaded by computer technology as ours, problems of accountability occur. More specifically, she identifies four barriers to ensure "accountability in a computerized society"[69] (Nissenbaum 1997). These four barriers are:
"1) the problem of many hands, (2) the problem of *bugs*, (3) blaming the computer, and (4) software ownership without liability" (Nissenbaum 1997: 41). Nissenbaum suggests that once a society is entrenched with computer technology, two related issues come to the fore: reliability and safety of technologies on the one hand and accountability in case of malfunctioning on the other. It is especially when systems are not highly reliable, that accountability becomes crucial. However, she argues that due to the before mentioned four barriers accountability in a computerized society is eroding, while reliability is still limited. In short, she argues that the more computerized a society gets, the harder accountability is to maintain and at the same time the need for accountability rises. Drawing on the concept of *moral blame* (Feinberg 1985, cited from Nissenbaum 1997), she analyses how these four barriers might threaten accountability in computerized societies.

The example of the *problem of many hands* serves as a good starting point because this problem if of special relevance for the analysis of epistemic social software, those socio-

[69] This is also the title of the essay (Nissenbaum 1997).

technical epistemic systems which form the center of this thesis. Nissenbaum argues that since most software solutions are not only produced in different settings, by a variety of people, but also often modular and integrated into other systems and applications, it becomes more and more difficult to target the source of error, to find the ones responsible and make them accountable for their actions and the consequences they have caused. By referring to one of the major examples of socio-technical epistemic systems used in this thesis, *Wikipedia*, the problem (but of course also the benefits) of *many hands* soon becomes obvious. The errors, for which you might want to make people responsible, could be severe misinformation or libel on the Wikipedia pages. Despite the constant reminder of the ease of revision to counter errors in Wikipedia, flamewars have not been rare, especially not on politically controversial entries. A tool that has been explicitly developed to improve the accountability of Wikipedia is the Wikiscanner, now labeled WikiWatcher.[70] With the help of this tool it is possible to track the institution behind the IP-numbers of the contributors. Due to the tracking of IP-numbers it was possible to show which organizations made which changes to Wikipedia. And of particular interest here are of course the changes that served the organizational, institutional or political interest of those making such changes in Wikipedia.

Using the computer as a scapegoat is another barrier Nissenbaum acknowledges (Nissenbaum 1997) and it should be seen as a way of shirking responsibility that should simply be avoided. The problem of *bugs,* however, may deserve more attention. *Bugs* are software errors that seem inevitable in programming. The problem with *bugs* as a threat to accountability is that they are clearly not intentionally inserted into software, but quite to the contrary a result of ignorance or neglect. The only way to amend for the threat of *bugs* to accountability lies in ensuring quality standards in software design that minimize the risk of major *bugs*. And this task must be considered an ongoing effort. The final barrier that Nissenbaum enlists is *ownership without liability* and it refers to current property right regulation over software. She criticizes that while traditionally, ownership not only comes with rights, but also with responsibilities, there is a recent trend in software copyright law to allow software companies to have rights without being responsible anymore.

[70] http://wikiscanner.virgil.gr/ [date of access: 18.02.2010]

To restore accountability, Nissenbaum offers three possible strategies. "Explicit standards of care" are guidelines of good practice or measure of excellence, that will be especially relevant to distinguish negligence from failures despite best efforts. With her request to "distinguish accountability from liability" Nissenbaum wants to stress that while "[...] appraisals of liability are grounded in the plight of a victim, [...] appraisals of accountability are grounded in the relationship of an agent to an outcome"(Nissenbaum 1997: 59). She argues that particularly with respect to the *problem of many hands*, differentiating between those two concepts might help to assure that no one who is accountable gets let off the hook too easily. With her final recommendation, she aims at enforcing "strict reliability and producer responsibility", because "[w]ell-articulated policies on liability would serve the practical purpose of protecting public interests against some of the risks of computer system failure which are further amplified by a reluctance on the part of producers and owners of systems-in-use to be accountable for them" (Nissenbaum 1997: 60f).

Freedom from Bias

Freedom from bias is a crucial goal for the development of any socio-technical system and much of the work in STS, as well as in feminist STS has focused on detecting and remedying bias in such systems. In their article on "Bias in Computer Systems" Friedman and Nissenbaum offer a taxonomy of biases that appears useful not only for future analyses of existing socio-technical system, but also as a guideline for the development of new tools (Friedman and Nissenbaum 1997).

First of all, what is bias in computer systems? The two authors use bias "[...] to refer to computer systems that *systematically* and *unfairly discriminate* against certain individuals or groups of individuals in favor of others" (Friedman and Nissenbaum 1997: 23). Two notions are crucial: for bias to occur it is not sufficient that people are treated unfairly, this treatment has to be systematic be called bias. Moreover, the systematic distortion alone is also not considered to be bias, the distortion must also lead to unfair results. (Friedman and Nissenbaum 1997) then identify three different categories of bias of relevance for computer systems: preexisting bias, technical bias and emergent bias.

Preexisting bias refers to "bias [which] has its roots in social institutions, practices, and attitudes" (Friedman and Nissenbaum 1997: 24). This is *the* type of bias that has stirred most analyses in STS. This form of bias is a classic example of all those societal injustices or personal prejudices that get inscribed into technology, be it intentionally or unintentionally.

Technical bias however is something different. This type of bias is not rooted in societal values, but rather arises within the process of technology design, when designers make technical decisions in certain ways and not in others, when they opt for one algorithm as opposed to another. The sources of technical bias that (Friedman and Nissenbaum 1997) list are limitations of computer tools, decontextualized algorithms, methods of randomization, and the biases that occur then human concepts have to be formalized to match the formats needed for computing.

Finally, the notion of *emergent bias* accounts for the fact that biases might occur later on through usage and appropriation of computer systems. Typically, such bias occurs when either the society in which the system is used changes, a process which Friedman and Nissenbaum describe as "new societal knowledge". The second reason for emergent bias has its roots in a mismatch between users and system design with respect to different expertise or values. This aspect is reminiscent of Madeleine Akrich's analyses of technologies that are used in contexts other than the ones where they have been developed (Akrich 1992).

Based on this taxonomy and the case studies they analyze, (Friedman and Nissenbaum 1997) draw several conclusions concerning the necessity to avoid bias in computer systems and define tasks for designers involved in the creation of computer systems. First of all, to avoid pre-existing bias, designers have to be aware of such pre-existing biases and be sensitive to what they possible inscribe into their systems. More precisely, the propose *rapid prototyping*, the *inclusion of different users groups* into the design process, *formative evaluation* and *field testing* as tasks that should minimize this type of bias. For technical and even more so for emergent bias, designers moreover have to envision they consequences and possible appropriations of their systems in different contexts. In order to achieve this, the authors propose the *anticipation of probable contexts of usage,* the *articulation of constraints* and the *possibility to take corrective action* once emergent bias becomes obvious.

They conclude their article with the following sentence, which I fully agree with: "Because biased computer systems are instruments of injustice – though admittedly, their degree of seriousness can vary considerably – we believe that *Freedom from bias* should be counted among the select set of criteria accord; to which the quality of systems in use in society should be judged" (Friedman and Nissenbaum 1997: 39). And despite the fact that *freedom from bias* can only be held out as an ideal, it is nonetheless the duty of designers to strive for this ideal.

3.3.4 Developing Methodologies: Two Examples

In the following two approaches are briefly portrayed that have explicitly tried to provide methodologies and guidelines for computer and information systems design, that account for insights from STS, computer ethics and other critical approaches. These two approaches are Batya Friedman's *Value Sensitive Design* as well as a the approach developed by Mary Flanagan, Daniel C. Howe and Helen Nissenbaum in the context of the project *Values at Play*.

Friedman 's Value Sensitive Design

The central role of Batya Friedman for the broader field of *Values in Design* has been indicated before. In this section, I want to depict her methodological approach, which she labels *Value Sensitive Design*. In a recent article on "Value Sensitive Design and Information Systems" Friedman and her colleagues describe their account in great detail and show how it takes insights *Computer Ethics*, as well as more empirical and technical fields, such as *Computer Supported Cooperative Work (CSCW), Social Informatics* and *Participatory Design* serious to develop methods and guidelines of technology design (Friedman, Kahn et al. 2006).[71] They define *Value Sensitive Design* as a "[...] theoretically grounded approach to the design of technology that accounts for human values in a principled and comprehensive manner throughout the design process. It employs an integrative and iterative tripartite methodology, consisting of conceptual, empirical, and technical investigations" (Friedman, Kahn et al. 2006: 348).

As noted before, Friedman and her colleagues define values quite broadly as that which people consider important in their lives. Clearly, such a broad concept leaves room for a variety of values of different degrees of abstractness. In their decidedly non-comprehensive list of values that may play a role in information system design, they include the following examples: human welfare, ownership and property, privacy,

---

[71] The description of their approach in this section is primarily based on this article of not noted otherwise.

freedom from bias, universal usability, trust, autonomy, informed consent, and accountability. With respect to systems design, they complement this list with those other values: courtesy, identity, calmness, and environmental sustainability (Friedman, Kahn et al. 2006).

Their methodology consists in an iterative integration of three parts: conceptual, empirical and technical investigations. *Conceptual investigations* encompass not only the identification of relevant values, but also the identification of different direct and indirect stakeholders. By including indirect stakeholders into the arena of analyses, they amend for the frequent neglect of non-users, of groups which may not be considered relevant but which are nonetheless affected by technologies (Wyatt 2005, Oudshoorn and Pinch 2005b). Relevant questions at this stage are concern the different stakeholders and the ways in which they are affected; the relative importance of different values as well as the trade-offs between conflicting values, etc.

The *empirical part* makes use of a diversity of quantitative and qualitative research methods from the social sciences to analyze how people actually conceive and prioritize different values, which role they play in the actual actions, etc. It is in this stage, that a *performative understanding of socio-technical systems* is taken serious, because in such an iterative, empirical methodology, usage and appropriation of technological artifacts can be observed and it can be analyzed whether the values intended in the design process were fulfilled, amended, subverted, etc.

The *technical investigations* as described by Friedman, Kahn et al. (Friedman, Kahn et al. 2006) consist of two parts. The first one focuses on the role values play in existing technologies and is in principle similar to analytic approaches portrayed in the sections before only with a decided focus on the technology itself. The second aspect is more interesting and innovative, since it concerns the "[...] proactive design of systems to support values identified in the conceptual investigation" (Friedman, Kahn et al. 2006: 352).

Friedman and her colleagues offer three case studies to show how their approach of *Value Sensitive Design* can be applied and indicate its points of originality. The case studies they offer concern cookies and informed consent in web browsers; the use of plasma displays to simulate windows in window-less offices; and UrbanSim, a simulation tool to predict patterns of urban development. Instead of depicting the details of these case studies, I conclude this description of *Value Sensitive Design* with the points of originality that they claim.[72]

First of all and crucially, *Value Sensitive Design* aims at being *proactive* in bringing forward the design of new value-sensitive artifacts instead of only analyzing existing technologies. With respect to values, they include a wide variety of moral values, usually not inherent in technology design. More precisely, they differentiate between *moral values and functional values*, such as usability and open up the possibility to weigh some values against others. Such value conflicts can not only occur between functional and moral values, but clearly also between different moral values, such as privacy versus security, accountability versus privacy, etc. They also take a useful stance in the debate of whether there is such a thing as *universal values*. They argue that the question of global or local values depends in the level of abstractness. For instance, the value of privacy in its most abstract form might be universally valid. However, what is meant by privacy, the extent to which is it needed varies in different settings.

Most crucially, they consider *Value Sensitive Design* be an *interactional* theory, in the sense that "[...] values are viewed neither as inscribed into technology (an endogenous theory), nor as simply transmitted by social forces (an exogenous theory). Rather, the interactional position holds that while the features or properties that people design into technologies more readily support certain values and hinder others, the technology's actual use depends on the goals of the people interacting with it" (Friedman, Kahn et al. 2006: 361). This stance clearly resonates with performative notions of the socio-technical described in the previous sections in that it allows for the possibility of appropriation, while considering design decisions to be relevant nonetheless.

In comparison to fields such as Computer Supported Cooperative Work (CSCW), they also have a broader set of applications in mind, which goes beyond the setting of workplaces. Moreover, they also broaden the scope of analysis by allowing not only for direct, but also for indirect stakeholder, and by doing provided a remedy for an overly exclusive focus on those stakeholders involved in the design and development of the artifacts. It is especially here, where power issues come into play, because as has been shown by various STS researchers, different stakeholder groups usually have different amounts of power. And finally, instead of remaining purely descriptive or analytic, they

[72] For information on these case studies confer Friedman, Kahn et al. (2006).

offer a clear methodology and practical suggestion and heuristics that can serve as a guideline for designing and developing value-sensitive artifacts.

Many of these *points of originality* resonate with approaches portrayed throughout this chapter. However, this should not come as a surprise given that *Value Sensitive Design* builds upon insights obtained in STS, Computer Ethics, Computer Supported Cooperative Work, Social Informatics and Participatory Design. Hence I would argue that it is not the single items that they list, which makes their approach unique. Rather it is the combination and their application to design practice that makes this approach so valuable.

Flanagan, Howe & Nissenbaum's Approach and "Values at Play"

Another interesting methodology in the field of *Values in Design* has been proposed by Mary Flanagan, Daniel C. Howe and Helen Nissenbaum (Flanagan, Howe et al. 2008). Since their proposal is partly informed by insights obtained from the project "Values at Play" (http://www.valuesatplay.org/), I shall briefly introduce it. According to its website, the project *Values at Play* "[...] was conceived with the intent of investigating how video game designers consciously and unconsciously embed social values into video games through narratives and game mechanics [...]" and aims at introducing "[...] designers to a systematic method for discovering, analyzing, and integrating values into their design work".[73] That is, the project not only attempts to make designers more conscious about the values they inscribe into the games they develop and the consequences this might have. It also delivers guidelines of how to promote certain values, such as equity, creativity or diversity. Hence, Flanagan and her colleagues develop their own pragmatic turn of insights obtained from STS to make them instructive not only for detecting values embodied in existing technology, but also to incorporate values willingly into technology. They propose that *social values* should be an integral part of the evaluation of technological artifacts in addition to other *functional values*, such as efficiency, reliability, usability, etc, by which technologies are usually assessed. The social values they have in mind are as diverse as privacy, trust, friendship, autonomy or transparency (Nissenbaum 2005). Their approach takes up the thread of *Value Sensitive Design* and combines it with others approaches, such as *Reflective Practice* (cf. Schön 1983, cited from Flanagan, Howe et al. 2008) and *Critical Technical*

[73] http://www.valuesatplay.org/?page_id=7 [date of access: 13.09.2009]

Practice (cf. Agre 1997, Dourish, Finlay et al. 2004, both cited from Flanagan, Howe et al. 2005; Flanagan, Howe et al. 2008).

They use another research project, *Rapunsel*, which focuses on teaching girls programming, as a case study to elucidate their claims. Indeed, it is instructive to take a closer look at the case study of *Rapunsel*, because the interaction of and possible conflicts between values become visible only within the design and development process. First of all in the process of game design, there is a diversity of sources of values: there are values in definition of project, which in the case of *Rapunsel* centered on *gender equity*. However, other values emerge in the course of design, for instance when implementing the reward systems guerdoning cooperative or competitive action. The other two types of values that play a role are more implicit and consist in the values and implicit assumption of the designers on the one hand and the users on the other. Especially when the implicit values between designers and users differ profoundly, appropriation and subversion can occur and it is here where empirical methods are needed to analyze the development of values in technologies.

In developing their account, they highlight some of the problems that might arise when trying to incorporate value considerations into the design process. These challenges are of *epistemological* as well as of *practical* nature. I shall address the epistemological challenges first. According to the authors, different forms of knowledge - or modes of knowing - need to be combined in value-sensitive design and development of technologies. The different modes they identify are the *philosophical, the empirical and the technical mode*.

The philosophical mode consists in reflecting upon the nature, the extension and intension of values, etc. Moreover, it has to offer some normative orientation in "[...] providing rationale or justification for commitments to particular values in a given device" (Nissenbaum 2005: xvii). The technical mode by contrast consists in bringing "[...] to bear state-of-the-art scientific knowledge and technical know-how on particular design specifications that realize given values in an overarching design project" (Nissenbaum 2005: xvii). And finally in the empirical mode two additional tasks have to be fulfilled. First it has to be empirically determined which values are relevant for different relevant groups in comparison to those that have been normatively set in the philosophical mode. Secondly, empirical methods have to be employed to assess whether a technology actually does embody the values it was intended to embody. The crucial question here is: did we succeed in embodying the desired values? Such an evaluation should rather be formative than summative in order to allow for the discovery and tracking of emergent bias (Friedman and Nissenbaum 1997). Similar to Friedman, Kahn et al. (Friedman, Kahn et al. 2006) they also propose an iterative process of integrating the philosophical, empirical and technical knowledge. To indicate the necessity of interaction and feedback between these different modes of inquiry, they use the metaphor of keeping "balls in the air" (Nissenbaum 2005 lxvii), resp. of keeping "balls in play" (Flanagan, Howe et al. 2008). And this request can also be read as a request for more and better interdisciplinary collaboration between those focusing on theoretical, empirical or technical issues.

The *practical challenge* concerns the lack or at least scarcity of concrete guidelines and methodologies for such an interaction. Their own guideline consists in iteratively making use of the following three activities.

> "1) Discovery, in which a list of values relevant to a project are compiled; 2) Translation, in which values are operationalized and implemented in material design features corresponding to these values; and 3) Verification, in which implementations are tested using variety of methods in order to ascertain whether designers' intentions have been met." (Flanagan, Howe et al. 2008: 347).

By applying the issue of *Values in Design* constructively to the field of game design, Flanagan and her colleagues propose an intentional inscription of values into design and they conclude their methodological proposal with the request for designers to "[...] embrace this dimension of their work, even if they are not always able to prevail against the tide of countervailing forces" (Flanagan, Howe et al. 2008: 350).

3.3.5 Conclusions

What conclusions can be drawn from this section on *Values in Design*? First of all, the fact that inscribing values into design is probably unavoidable is not a reason to despair. If evil designers can inscribe their disvalues into technologies, then those designers who aim at making our societies more just can also use technologies to reach their goals. However, a performative caveat has to me made: We cannot fully control the consequences of our (technical) decisions. Even the best intentions and the most rigid testing cannot safeguard us from negative consequences of our decisions. Such an acknowledgement of the entangled and complex nature of the socio-technical environment we live in cannot and should not be equated with shirking off responsibility. Despite the difficulty of the task, the constant need for cross-checking, reformulations and the lack of security that the decision we make will turn out well, we have to take stances. Resonant of the requests by feminist STS scholars I would like to conclude that in order to make our worlds more livable, we have to take active part in the design of our socio-technical environment, we have to make sure that frames get expanded and the necessary cuts are made in accountable ways.

In the concluding remarks of the introduction to her seminal edition Batya Friedman urges her community to "[...] embrace the value-sensitive design as part of the culture of computer science" and argues that this step may lead to "[...] the conceptualizations needed to identify shortcomings in current designs and to seek remedies which prompt human well-being. It moves us toward the language needed to discuss the often immense social consequences of our work with the public at large. And it moves us toward holding out value-sensitive design as a criterion – along with the traditional criteria of reliability, efficiency, and correctness – by which systems may be judged poor and designers negligent. As with the traditional criteria, we need not require perfection, but commitment [...]"(Friedman 1997b 13).

I would like to add that we should not leave all the work for the computer scientists, but rather join them in this endeavor. Since no one can do all the tasks necessary in critical technology design on one's own, this is also a request for more interdisciplinary collaboration. For the topics of this thesis, this collaboration especially refers to a desirable collaboration between science and technology scholars, epistemologists and software designers.

4 Summary and Conclusions of Part 1

In the previous chapters I have introduced the topic and the empirical basis of this thesis: *epistemic social software defined as socio-technical epistemic systems in which multiple human and non-human agents interact to create epistemic products*.

In Chapter 2 I gave a brief and very cursory introduction into the ill-defined success of Web2.0, also referred to as social web, social media or social software. These last three terms refer to a crucial characteristic of this phenomenon, which formed the empirical starting point of this thesis: its sociality. In social software applications usually multiple users interact with a technological infrastructure or with each other through this infrastructure for various purposes. The focus of this thesis was set on social software application whose purpose is decidedly epistemic. Because of this focus on epistemic results I have dubbed these applications *epistemic social software* to distinguish them from other social software applications with different foci. While arguing for the analytic utility of the distinction between epistemic social software and other social software, I have nonetheless acknowledged that this distinction depends on usage and may indeed be permeable and fuzzy.

I have focused on four examples of epistemic social software: the Web-encyclopedia Wikipedia, the social bookmarking services Delicious, the open nnovation platform InnoCentive and finally recommender systems as a more embedded form of epistemic social software. I have chosen these examples not only for their explicit epistemic focus, but also because they are characterized by very *different forms of epistemic sociality*. I return to these examples in more detail in Part 3 of this thesis and argue that different socio-epistemic mechanisms are employed and combined in each of these systems.

In the 3rd Chapter, I have argued that instead of referring to *tools* or *applications*, epistemic social software should better be understood as socio-technical epistemic systems consisting of multiple human and non-human agents who interact for various epistemic purposes. To support such an understanding of epistemic social software as socio-technical system comprising of entangled, heterogeneous agents I took a brief detour through the field of Science and Technology Studies and some neighbouring disciplines. The main insights I took from the various approaches portrayed can be clustered around the following topics: the *entanglement of the social, the technical and* *the epistemic*; the *relationship between human and non-human actors*, the *relationship between values and technologies* as well as a *performative understanding of epistemic practices and systems.*

I have argued that *epistemic social software* is a term to denote systems in which social, technical and epistemic aspects are deeply entangled. More specifically, epistemic social software, such as Wikipedia, Delicious, InnoCentive or Recommender Systems are socio-technical epistemic systems in which human and non-human agents interact on multiple levels and in different forms. Conceiving the technological components of these systems as non-human agents emphasizes the crucial and active role they play within the systems. However, this emphasis should not conceal the crucial differences between human and non-human agents. Since the attribution of *responsibility and accountability* is bound to *intentional agents*, a distinction between human and non-human agents has to be kept up. Within entangled socio-technical epistemic systems accountability and responsibility may be hard to track. Indeed, it has been argued that the necessity and the difficulty to locate accountability rise simultaneously in computerized societies.

During the development of socio-technical epistemic systems different values are necessarily inscribed into them. Such a *valuing of systems* in unavoidable, because each decision in developing systems in based on certain epistemic and societal background assumptions. These inscriptions may be unavoidable, but they need to be monitored because such embedded values can retroact on societal values in general by either reinforcing or degrading them.

Finally, socio-technical epistemic systems are not closed systems. Instead they develop and change through practices. Despite the relevance of programming decisions, users can to varying degrees appropriate and change systems through their own practices. Indeed, if technological artifacts can be appropriated, changed and subverted through usage, this is all the more true for socio-technical epistemic systems. The performativity of these systems is much more profound*: epistemic social software exists only through usage and practice*. Without the processes and interactions between human and non-human agents within these systems, epistemic social software, such as Wikipedia, Delicious, InnoCentive or Recommender Systems cease to exist. Hence, this fundamental process-dependence has to be acknowledged in any analysis of such socio-technical epistemic systems. It may be harder to address and tackle such dynamic, performative and entangled systems, more difficult to locate accountability and responsibility within them. But there is no way in avoiding this complexity.

One of the primary consequences of these insights is that *transparency* should be a crucial value in assessing socio-technical epistemic systems. Transparency is needed to understand how systems function, to comprehend their strengths and weaknesses. It is needed to uncover and understand which values are promoted or undermined in such systems. Further, transparency is necessary to locate accountability and responsibility within such systems, especially in cases of misconduct and error. But transparency is not only needed to attribute accountability and responsibility to others. It is also needed to act responsibly oneself. Only if I understand how socio-technical epistemic systems function, I can act as a *responsible knower*. Only if I understand the modes of operation, the benefits and shortcomings of different systems, can I become a responsible, accountable and empowered knower myself. The knowers in socio-technical epistemic systems are the crucial figures in the next part of this thesis, in which I introduce *social epistemology as the philosophical discipline addressing the social nature of knowledge*.

PART 2

5 Social Epistemology

5.1 From Socio-Technical Epistemic Systems to Social Epistemology

In the previous chapters I have introduced the notion of epistemic social software and argued that understanding them as socio-technical epistemic systems consisting of multiple epistemic agents in interaction provides a more fruitful framework for analyzing them than simply referring to them as *tools* or *applications*. Observing the different socio-epistemic mechanisms employed in such systems, the relationships and interactions between different epistemic agents, I became interested in the ways in which knowledge in general can be conceptualized a result of processes which are social, technical and epistemic at the same time. While the technical has been the starting point and core of the previous chapters, I shift the focus now and take a closer look at the relationship between the social and the epistemic. A key characteristic of the systems portrayed before is the utilization of multiple agents for epistemic purposes: several people are engaged in the creation or distribution of knowledge in various ways. Shedding some light on how these agents interact, communicate, are coordinated and orchestrated within these socio-epistemic processes is the central task of the following chapters. The field I have chosen as a theoretical framework to analyze these processes and the roles of different epistemic agents within them is *social epistemolog*y.

5.2 Addressing the Social Dimensions of Knowledge

> “Social epistemology is the study of the social dimensions of knowledge or information. There is little consensus, however, on what the term "knowledge" comprehends, what is the scope of the "social", or what the style or purpose of the study should be.” (Goldman 2006).

Social epistemology is the philosophical discipline exploring the ways and the extent to which knowledge and epistemic practices are social. As a term, social epistemology often refers to a quite specific field of discourse which “dates from the 1980ies, is primarily a philosophical enterprise, and has its roots in Anglo-American epistemology, in feminist theory, as well as in the philosophy of science" (Kusch to appear: 1). This is what Martin Kusch, one of the philosophers whose approach is depicted below in some detail, considers to be the narrow understanding of social epistemology - and it forms the theoretic core of this thesis.

Talking about a *narrow* understanding of social epistemology implies that *broader* understandings of social epistemology are possible. And indeed the ways and the extent to which knowledge is social have been addressed within philosophy as well as within other disciplines in numerous ways. Throughout the history of philosophical thought philosophers have addressed different social dimensions of knowledge. In encyclopedic articles and introductions to social epistemology (e.g. Goldman 2006, Goldman 2009a, Kusch to appear) reference is given to a diversity of predecessors. Plato's explorations of how laypersons can determine expertise in *Charmides* is often used as one of the earliest examples (cf. Goldman 2006).

The label under which the social nature of knowledge has been primarily discussed within Western philosophy is *testimony*. Testimony ususally refers to the "assertion of a declarative sentence by a speaker to a hearer or to an audience" (Adler 2006) and it is considered to be the fourth route to knowledge besides memory, perception and inference. One of the main *epistemological questions* around testimony concerns the *status of knowledge* acquired through testimony: Does knowledge received through the words of others have a different quality than knowledge obtained via one's own cognitive resources, i.e. perception, memory or inference? Given that many philosophers nowadays acknowledge that we acquire far more knowledge through testimony than through our own onboard resources, indeed that some of them even state that (almost) everything we know depends on interaction with others, it might come as a surprise that knowledge obtained from testimony has long been considered to be of secondary status. Knowledge received from testimony was considered to be less valuable, less reliable than knowledge obtained from memory, perception or inference. Indeed, sometimes knowledge obtained via testimony was denied the status of knowledge altogether. The reason for this depreciation of testimony is often seen in the epistemic individualism, i.e. in an individualistic bias within Western philosophy. This inferior epistemic status of testimony combined with a focus on one's own onboard cognitive resources is most clearly stated in Locke's famous dictum that "[w]hat in them was science is in us but opiniatrety." (Locke 1961: 58, quoted from Coady 1992: 14).[74]

Beyond classical epistemology other predecessors of social epistemology exist within philosophy. Martin Kusch has argued that if one understands social epistemology broadly as "[...] all systematic reflection on the social dimension or nature of cognitive achievements such as knowledge, true belief, justified belief, understanding, or wisdom" (Kusch to appear: 1), then many contributions from Marxism, Critical Theory or Hermeneutics also qualify as socio-epistemological.

Considerable overlap exists between social epistemology and many feminist epistemologies. Indeed, according to Grasswick "the significant body of work of feminist social epistemologists has provided key theoretical resources for understanding the social dimensions of knowing" (Grasswick 2006: 1). She argues that feminist epistemologists have been particularly important in revealing and criticizing the individualism of contemporary (analytic) philosophy and in developing alternative models of knowers as social beings that are *situated in different contexts* (e.g. Code 1991, Code 2001, Harding 1991), that are in *interaction with each other* (Alcoff 2001, Scheman 2001) and that *depend on the communities they are part of* (Longino 2002c, Nelson 1993). Moreover, several feminist epistemologists have developed social models of knowledge and objectivity (Haraway 1996, Harding 2003, Longino 2002c, and stressed the relationship between epistemology and ethics (e.g. Fricker 1998; Fricker 2007).

Many social epistemologists have also acknowledged the relevance of different sociological and historical approaches for understanding the ways in which knowledge is social. Within sociology, particular emphasis has been placed on the sociology of knowledge (e.g. Mannheim 1936), the sociology of science (e.g. Merton 1973), the sociology of scientific knowledge (e.g. Barnes and Bloor 1982, Collins 1981b; Collins 1985), as well as different approaches within the field of Science and Technology Studies, such as the works of Bruno Latour (Latour and Woolgar 1986, Latour 1987) or Karin Knorr Cetina (Knorr-Cetina 1984; Knorr-Cetina 1999). Among the historical approaches, some of the most received works are Thomas Kuhn's "The Structure of

[74] I return to the topic of testimony in the next chapter when outlining central socio-epistemic mechanisms.

Scientific Revolutions" (Kuhn 1962/1970); Steven Shapin's "A Social History of Truth" (Shapin 1994) as well as "Leviathan and the Air-Pump" (Shapin and Schaffer 1985); Michel Foucault's analyses on the relationship between knowledge and power (e.g.Foucault 1970; Foucault 1971; Foucault 1980), and finally the works that fall under the label of *historical epistemology*, most notably the works of Ian Hacking (e.g. Hacking 1992a; Hacking 2004), Lorraine Daston and Peter Galison (e.g Daston 1993; Daston and Galison 2007).

Finally, different social epistemologists have proposed directions into which social epistemology should be heading. Kusch has proposed a differentiation into five different directions and listed philosophers whose work could be instrumental for the future of social epistemology (Kusch to appear): analytic social epistemology (with reference to the works of Donald Davidson (Davidson 1991/2001), genealogical social epistemology (Craig 1990), historical social epistemology (Shapin 1994), political social epistemology (Harding 1991) and naturalistic social epistemology (Hardwig 1985; Hardwig 1991). Alvin Goldman, one of the most central figures in social epistemology, finally has proposed a new perspective for the future of social epistemology which he labels *systems-oriented social epistemology* understood as "a flexible form of epistemological consequentialism that evaluates social epistemic systems in terms of their impact on epistemic outcomes" (Goldman 2009c).

5.3 Social Epistemology: Delineating the Field

Clearly, this multitude of approaches allows for different ways of apprehending the sociality of knowledge, different ways of cutting the field, different ways of asking different questions. In the previous chapter, we have learned that objects of analysis are not given, but made by decisions. Each decision to include or ignore something comes with consequences and one can be held accountable for one's decisions. Nonetheless, decisions – and omissions - are unavoidable.

I have chosen what Kusch considers to be the narrow conception of social epistemology as the theoretical nucleus of this thesis: an inquiry which "dates from the 1980ies, is primarily a philosophical enterprise, and has its roots in Anglo-American epistemology, in feminist theory, as well as in the philosophy of science (Kusch to appear: 1).

However, neither is this core *indivisible* nor completely *unrelated* to other fields of reasoning investigating the social nature of knowledge outlined before.

More specifically, *within* this narrow understanding of social epistemology there are several major lines of demarcation and numerous classifications of social epistemologies have been proposed (e.g. Goldman 2009a; Goldman 2009b; Goldman 2009c, Goldman 2006, Kusch to appear, Kusch 2002, Quinton 2004, Kitcher 1994, Fuller 2004, Kornblith 1994, Schmitt 1994a). Instead of listing and explaining all the different labels that have been proposed to classify social epistemologies beforehand (e.g. critical versus analytical social epistemology, classical versus anti-classical approaches, etc.), I have decided to address these differences in next chapter *after* having introduced five major social epistemologies. Some of the most important differences concern the adherence versus rejection of the classical analytic definition of knowledge as justified true belief; the question of whether social epistemology is complementary to individual epistemology or rather its successor; as well as the question what the goals of social epistemology should be. My hope is that different stances towards these issues will become clearer through my portrayal of the different socio-epistemological approaches.

With respect to the exoteric relationships, different approaches beyond classic epistemology are often used as background foils for the development of social epistemologies. That is, many social epistemologists sharpen their approaches with reference to and often *against* various theories from the sociology of scientific knowledge, against feminist standpoint theories, etc. Sometimes theories on the sociality of knowledge that are not rooted in an analytic philosophical tradition have received extremely critical or even hostile reactions from certain social epistemologists (e.g. Goldman 2003, Kitcher 1994). Other social epistemologists, however, have stressed the importance of alternative philosophical positions as well as sociological and historical analyses for the development of their social epistemologies (e.g. Kusch 2002, Longino 2002c). In any case, an understanding of the central arguments within social epistemology even in its most narrow understanding is impossible without seeing the broader field of discourse in which these arguments are embedded. Thus, although a narrow conception of social epistemology as the mainly analytical, Anglo-American philosophical disciplines that has emerged in the late 1980ies is the core, it is rather the *field of discourse around this core that forms the theoretical framework of this thesis.*

In the following I outline five major socio-epistemological theories: Steve Fuller's *Social Epistemology*, Alvin Goldman's *Veristic Social Epistemology*, Miriam Solomon's *Social Empiricism*, Martin Kusch's *Communitarian Epistemology* and Helen Longino's *Critical Contextual Empiricism*. I have decided to depict these approaches in some detail, because of their importance for the field and their comprehensiveness. All of them have been developed in monographs and describe themselves either explicitly as social epistemologies or situate themselves in the discourse field of social epistemology.

Numerous other authors have addressed certain social aspects or dimension of knowledge in articles, but to my knowledge no other comprehensive social epistemology has been proposed in monographic length. To give an overview over the field, I portray the different theories in some detail in this chapter in a rather neutral way, i.e. without critical assessment. I offer such a critical evaluation as well as an analysis of the merits of the different social epistemologies for the questions of this thesis in the next chapter. Instead of critically assessing the approaches *individually* directly after portraying them, I use my critique to delineate the major topics, debates and controversies within the field of social epistemology. In this analysis other theoreticians, who have made important contributions to specific social aspects of knowledge and epistemic practices only are also included.

My analyses in the next chapter also serve as the basis for the development of my own approach, which will be outlined in the third part of this thesis. I offer a new, systematic perspective for understanding and analyzing socio-epistemic practices on the Web and beyond that is based on my critical reception of the different social epistemologies to be portrayed. Before heading into the details of the five before-mentioned comprehensive social epistemologies, I take a brief look at the coining of the term 'social epistemology' – a dual occurrence which is of interest for this thesis.

5.3.1 Social Epistemology: The Coining of the Term

Social Epistemology in the Library Science

It was in 1987 that the term *social epistemology* was first introduced into the philosophical debate by a special issue of the journal *Synthese*. But the term had already been used in the early fifties in a quite different context. Despite its philosophical ring and the subsequent philosophical occupation of it, the term *social epistemology* was initially coined in the library sciences by Mary Egan and Jesse Shera. They introduced the term in trying to formulate an epistemological foundation for library science (Egan and Shera 1949; Egan and Shera 1952). In a paper on "Classification as the Basis of Bibliographic Organization" Egan and Shera write:

> "Even a cursory examination of the history of the classification of the sciences emphasized the extent to which any attempt to organize knowledge is conditioned by the social epistemology of the age in which it was produced. This dependence of classification theory upon the state of the sociology of knowledge will doubtless be even more strongly confirmed in the future." (Shera 1950: 82).

Although the tasks of librarians in classifying and providing information probably exists as long as knowledge was recorded, the field of library science went through process of professionalization at the end of 19th and the beginning of the 20th century (Zandonade 2004). And while the usage of the term *social epistemology* in the quote above seems to refer rather to a *descriptive and retrospective analysis* of the ways in which the social environment has shaped the organization of knowledge through classifications, there were also more *normative and prescriptive concerns* related to the field of library science. Egan and Shera therefore intended social epistemology to be "[...] a new discipline [...] that will provide a framework for the effective investigation of the whole complex problem of the intellectual processes of society" (Egan and Shera 1952: 132, cited from Fallis 2006: 476). Information science according to Shera should focus on the "production, flow, integration, and consumption of all forms of communicated thought throughout the entire *social* fabric" (Shera 1970: 86, quoted from Fallis 2006: 482f), which is why a broader socio-epistemological perspective was required. Social epistemology understood in those terms was meant to provide *empirical* data about the sociality of epistemic content (such as classification) and epistemic practices. And the normative aspiration was to feed this empirical information back into library science in order to *improve* epistemic practices.

For the purpose of this thesis, these roots of social epistemology in the field of information science, or rather its predecessor library science, is clearly striking. Given my aim of analyzing the epistemological relevance of new socio-technical systems, often referred to as Web2.0, my thesis can also be regarded as an attempt to resuscitate this relationship between information science and epistemology as two fields, which are crucially related to knowledge and epistemic practices, albeit in different ways. In arguing for an *applied social epistemology*, that not only takes into account actual epistemic practices of humans for its theorizing, but also seeks to make recommendations of how to improve socio-technical systems for epistemic purposes, I do attempt to pick up the thread where Egan and Shera's have left it off.

Social Epistemology: Emergence as a Philosophical Discipline

Within philosophy, there are two – albeit related - funding acts that mark the beginning of the field of social epistemology as conceived in this thesis, one accomplished by Frederick F. Schmitt, the other by Steve Fuller. In 1987, Frederick F. Schmitt's edited a special issue of the journal *Synthese,* in which several authors emphasize social aspects of knowledge, cognition and belief (Schmitt 1987). This special issue was entitled "Social Epistemology" and included contributions on social features of cognition (Kornblith 1987), social standards (Cohen 1987), justification and sociality (Schmitt 1987b), personal and social knowledge (Lehrer 1987), collective belief (Gilbert 1987). Moreover, it featured an article by Alvin Goldman on social epistemics (Goldman 1987) and one by Steve Fuller (Fuller 1987) on the regulation of scientific knowledge creation.

These early articles by Fuller and Goldman already contain the seeds of the two social epistemologies later to be developed in greater detail by both authors. Goldman and Fuller moreover, emerged to be two central figures in the field of social epistemology. Not only have both of them provided monographs and numerous articles on social epistemology themselves. Both of them have also been long lasting editors of the two major journals devoted to social epistemology. While Fuller established his journal *Social Epistemology* already in the following year, 1988, Goldman became the editor of *Episteme* one year after its founding in 2005. Goldman's *Veristic Social Epistemology* is depicted in more detail later on, and I continue with a summary of Steve Fuller's social epistemology right below.

In 1994, Schmitt also edited a seminal book on social epistemology entitled "Socializing Epistemology: The Social Dimensions of Knowledge" (Schmitt 1994d). In this anthology, many social epistemologists, whose work is portrayed in this thesis are already gathered. In addition to the before mentioned authors, this refers in particular to Helen Longino, Miriam Solomon, Philip Kitcher and C.A.J. Coady.

5.4 Major Social Epistemologies

5.4.1 Steve Fuller's Social Epistemology

In his article in the 1987 Synthese special issue on social epistemology (Fuller 1987), Steve Fuller introduces the term "social epistemology" - still in quotation marks – as the label for his account of the sociality of knowledge and science. In 1987 Fuller also funded a journal and a year later published a book with the same title "Social Epistemology" (Fuller 1988), in which he depicts his approach in greater length. For these reasons and the fact that he has been the first philosopher using the label *social epistemology* for a distinct approach, his approach has to be depicted, although it is not of major relevance for the rest of this thesis.

From a theoretical and a less institutional - or rather institutionalizing - point of view an important contribution of Fuller consist in the identification of certain blind spots or shortcomings of many accounts of science within philosophy of science and epistemology as well as in the social studies of science (Fuller 2004, Fuller 2006, Fuller 1994). In particular, he has emphasized the neglect of the political, institutional and organizational, economic contexts of philosophical accounts of science and the lack of normative concerns in their sociological counterparts. Fuller further insists that, both the philosophers and the sociologists should take a more critical stance towards science instead of assuming that science has to be measured by its own standards and that it is in principle working just fine.

Fullers' own approach focuses explicitly on taking such a critical stance and aims at *changing science*. His social epistemology in principle is an epistemologically informed science policy that sets on to change science by changing its social structures and those it is embedded in. The hope that is related to this macro-approach is that by changing the *structure* of science, the *content* of science, the kinds and types of knowledge that get produced, could be changed accordingly. Thus, Fuller argues against an "invisible hand" model of scientific rationality, according to which science is at best left alone in regulating itself. Instead he installs the social epistemologists as a science policy maker, who exerts influence on the way science is conducted and on the kinds of knowledge that get produced.[75]

It is in the before mentioned special issue of the journal *Synthese* where Steve Fuller offers a first sketch of his social epistemology. He considers its major question to be the following:

> "How should the pursuit of knowledge be organized, given that under normal circumstances knowledge is pursued by many human beings, each working on a more or less well defined body of knowledge and each equipped with roughly the same imperfect cognitive capacities, albeit with varying degrees of access to one another's activities?"(Fuller 1987: 145)

Fuller delineates his social epistemology as *normative* and *naturalist* social epistemology and conceptualizes the social epistemologist as an "ideal epistemic policy maker" (Fuller 1987: 145). His social epistemology is *normative* in that it aims at arriving at an *optimal distribution of cognitive labor*, an interest that he shares with many other social epistemologists, such as Miriam Solomon (Solomon 2001), Philip Kitcher (Kitcher 1993) and Alvin Goldman (Goldman 2003). Moreover, the theory is also normative in the sense of *changing the content of science* by changing the social structure of science. Fuller asserts to be a *naturalist* by focusing on the *normal contexts* in which knowledge is produced. And since these contexts are mostly social contexts and not the Cartesian lonesome thinker, he concludes that his epistemology has to account for this sociality.

Let's take a look at *how*, *by what means* and *to what ends* Fuller thinks science should be changed. To begin with, Fuller argues that while philosophers and sociologists of science might differ in whether it is *methodology or interests* that steer science, both camps seem to take for granted that science runs quite well either way (Fuller 1992: 392). That is not the case, Fuller announces and argues that there is a "[...] big difference between claiming that science works well enough to sustain itself and claiming that it works optimally toward a desired outcome" (Fuller 1992: 394). Two interpretations of this quote are possible. Fuller may refer to the prospect of *improving the efficiency of science*. The other reading concerns the dependence of this evaluation on the *desired outcomes*. Let's start with how to improve efficiency.

[75] In this reference to science policy as well as in the intended use of historical data to develop guidelines for it, Fuller's approach is closest to Miriam Solomon's *Social Empiricism* (Solomon 2001), which is portrayed below.

How could we ensure that science lives up to its full potential, that is does not merely function well enough, but well? In principle, Fuller argues that the best way to improve science is by applying scientific reasoning to the evaluation and subsequent improvement of scientific practice itself. More specifically, Fuller argues for an *experimental analysis and correction of scientific practice* in the spirit of Taylor's *Scientific Management* (Fuller 1992: 413ff). He argues that "[...] there is a pressing need to examine not merely how science works but whether science is working as well as it could, especially given the ever-changing and ever-more-important roles that science plays in society (Fuller 1992: 395). The first step in *improving science* according to Fuller has to consist in a *decomposition* of the different scientific tasks and sub-tasks instead of treating science as an "organic skill" over which only the scientists themselves have complete authority (Fuller 1992: 414). In treating scientific practices in the same way in which Taylor treats his workers' coal shoveling practices, namely by decomposing it into more minute sub-tasks and improving them separately, such a holistic view of science is overcome and the possibility to improve scientific efficiency is opened up.

The social epistemologist would be the external observer of science, who due to having analyzed scientific practices is in the best position to *improve* science by improving the conduction of individual subtasks and their distribution over multiple people. Since the social epistemologist has analyzed the minute tasks, he is in a better position than the scientists themselves to make science more efficient. Therefore, the social epistemologist revokes the scientists' authority of how science should best be conducted in much the same way in which Taylor revoked the coal workers' authority over how coal is best shoveled.

The second change that Fuller's social epistemology imposes on science does not refer to its *practices*, but to its *goals*. In contrast to many philosophers of science, Fuller argues for a science in which "[...] the "ends of science" [...] are not given by science itself but by something else to which science is held accountable" (Fuller 1992: 395). Science should be more accountable to the public, which implies that science and the public need to get back into a dialogue. This means that on the one hand, the public needs to be made more scientifically literate (Fuller 1992: 396) and on the other hand, scientists need to be made more publicly accountable. While the former seems to evoke some form of science education, similar to the "public understanding of science" program (Bodmer 1985, Wynne 1995), the latter seems to refer to some "personnel development measure" in which scientists are taught "how to deal with people" (Fuller 1992: 397).

Moreover, once science is aligned with societal needs, science has to be evaluated with respect to the fulfillment of these societal needs. This implies that one cannot change the goals on the run, but that one has to stick to them – unless otherwise consented upon. Unfortunately, Fuller argues that "a crucial way in which a discipline maintains its status as "science" is by manipulating the historical record so that it appears to be the inevitable outcome of the course of inquiry up to that point" (Fuller 1987: 145). This would have to be changed according to Fuller's social epistemology: science has to be accountable to its *original standards* and goals and not to the ever-changing *a posteriori* goals scientists fabricate to justify their research at hindsight.

I have asserted before that Fuller's social epistemology is of little relevance for this thesis. Given the historical importance of Fuller's social epistemology, this negative assessment may need some vindication. Although I have announced that I remain rather descriptive in this chapter and postpone a critical assessment to the next chapter, I include a brief critical assessment of Fuller's account already at this point, precisely because I do not return to it later on. I appreciate Fuller's emphasis on the institutional and political contexts of science and his goals to make science address societal needs and to make it more accountable to the public. The relationship between societal needs and scientific goals has been debated hotly in philosophy of science. While many feminist epistemologists and philosophers of science have argued for an alignment of science to match the most pressing societal needs (e.g. Kourany 2003), others have defended the non-instrumental value of science (e.g. Ziman 2003). A constructive approach of how to overcome this dichotomization of "pure" research and applied or instrumental research has been delivered by Martin Carrier (Carrier 2004; Carrier 2008), who has also argued for a concept of a *science in the service of the public.*[76]

I also consider Fuller's suggestion that social epistemology should be a multi- or interdisciplinary endeavor that might play a crucial role for science policy interesting and valid. Despite this more general consilience, I am less convinced by Fuller's own

[76] Carrier referred to the possibility of such a *science in the service of the public* in a talk delivered in the Spring School on "Science and Values", which took place at the University of Bielefeld in April 2008).

theoretical developments. Amidst an abundance of references to a variety of different lines of thought on the sociality of knowledge and science, I find it quite difficult to identify Fuller's main arguments beyond rather general recommendations concerning science policy outlined above (e.g. Fuller 1988). I do think that Fuller's account is broad in the variety of societal factors he acknowledges as well as in the reference he makes to various disciplines (most notably to the field of science and technology studies (Fuller 2006). However, his analysis of the various kinds of social processes and mechanism, the various types of epistemic sociality that play a role *within* science is surprisingly thin. By focusing on the distribution of cognitive labor and reducing social epistemology to science policy, Fuller ignores other important aspects of what it means for knowledge to be social. Such neglected aspects include collaborative practices in the sciences, the different social forms of knowledge creation, etc. Therefore, I would argue that restricting social epistemology to science policy clearly is premature and forecloses a more thorough analysis of the different ways in which knowledge is social.

Another point of critique concerns his reliance on Taylorism as a model for science management. While I agree that it might not be inherently problematic to use a mechanism from economy to amend scientific practices, I find it surprising that Fuller chooses *Scientific Management* from all available possibilities. After all, Scientific Management has not really been the most sustainable success story in the history of economics. As even Fuller acknowledges, *Scientific Management* failed due to the worker's protests and resistance (Fuller 1992: 422). Why should scientists react any different to such an incapacitating and disempowering process? Surely, simply referring to advocacy advertising or incentives does not seem to be too convincing.

5.4.2 Alvin Goldman's Veritistic Social Epistemology

<u>Introduction</u>

Alvin Goldman is one of the most prominent social epistemologists. He has clearly shaped the field of analytic social epistemology profoundly not only through his long-lasting editorship of the journal *Episteme*. His own publications, especially his widely received book "Knowledge in a Social World" have also inspired a lot of socio-epistemological work. In this book, Goldman develops a *veritistic* social epistemology and applies it to a variety of different cases in science, law, politics and education. Two aspects are of particular interest for this thesis. Goldman not only emphasizes the necessity of applying epistemology to real-life situations. He also develops his social epistemology as a "philosophy for the information age" (Goldman 2003: back cover). Given the importance of Goldman's approach for the whole field of social epistemology and the similarity of its intended focus with the focus of this thesis, I portray Goldman's social epistemology in some detail using his book "Knowledge in a Social World" as the primary basis for my portrayal.

"Knowledge in a Social World" is divided into three parts. In part one, Goldman lays the foundations of his veritistic social epistemology. The goal of this veritistic social epistemology is to evaluate social practices with respect to their conduciveness or detriment to the attainment of truth. In part two he applies his veritistic social epistemology to four different types of generic social practices: testimony, argumentation, information technology and speech regulation. In the third part, he investigates special domains in which those social practices play a role. These special domains are science, law democracy and education.

In the following I outline several aspects of his approach in more detail. At first, I portray some of the fundaments of Goldman's veritism. For the purpose of this thesis emphasis is then laid on Goldman's account of testimony, his recommendations for veritistically beneficial forms of argumentation, as well as the relevance of information and communication technologies (ICT) for a veritistic social epistemology. Regarding his analyses of specific societal domains, I focus especially on the domain of science.

Key Features

Complementarity

According to Martin Kusch (Kusch 2002), Goldman is the prime representative of the *complementary approach in social epistemology*, which is characterized by a clear distinction between social and individual aspects of knowledge. For Goldman, social epistemology is meant to account for the social dimensions of knowledge, whereas there is a distinct individual dimension of knowledge that can best be described by classical epistemological theories and theories from the cognitive sciences. Individual epistemology would focus on beneficial psychological and cognitive processes within epistemic agents whereas social epistemology would focus on social and institutional contexts of epistemic practices, such as interactions between agents to identify and evaluate social and communicational processes. This complementary approach is also reflected in the succession of Goldman's major publications: only after having investigated the respective relevance of psychology and epistemology for individual knowledge in great detail in his book "Epistemology and Cognition" (Goldman 1986), his attempts to socialize classical epistemology culminate in "Knowledge in a Social World" (Goldman 2003) more than a decade later.[77]

Normativity

Moreover, Goldman emphasizes the *normative function of epistemology*. That is he stresses that in opposition to sociological theories of knowledge epistemology cannot remain purely descriptive. Such normativity in Goldman's approach is twofold. On the one hand, epistemology should deliver criteria by which knowledge can be distinguished from mere belief or opinion. On the other hand, normativity is also understood in a prescriptive sense of identifying practices that are epistemically valuable and that should therefore be fostered. The goal of epistemology according to Goldman is to provide normative standards for epistemic practices. Epistemology should "seek to identify and assess processes, methods and practices in terms of their contribution – positive or negative – to the production of true belief" (Goldman 2006).

Veritism

The most important characteristic of Goldman's approach however concerns the role of *truth*. That the attainment or approximation of truth looms large in Goldman's account becomes obvious not only from the previous quote. He also labels his whole approach *veritistic* social epistemology. To ground his veritism, Goldman argues that information seeking is a crucial activity in everyday life – be it for practical reasons or for mere curiosity. Moreover, for the most part we seek *true information*, or at least approximately true information. Therefore a proper epistemology – be it individual or social – should better come up with some help for the attainment of truth.

Goldman defines knowledge in the "weak" sense of (merely) "true belief", i.e. he is concerned with the truth value of a proposition and not with its justification. And on the

[77] Although Goldman initially intended to synthesize social and individual aspects of knowledge in one book, they were nonetheless conceived to be different processes. And while the individual epistemology should be related to cognitive psychology, the empirical basis for a social epistemology was to be found rather in social and political science (Goldman 2003: ix).

first almost 100 pages of his book, Goldman seeks to defend the importance of truth in science and everyday life against a variety of enemies. At first he notes that some approaches that so far have been labeled social epistemology, namely Steve Fuller's "Social Epistemology" (Fuller 1988) and Shapin's "Social History of Truth" (Shapin 1994) do not deserve the label *epistemology* and should rather be tagged "social doxology", because they cannot and do not differentiate between knowledge and opinion (Goldman 2003: 7). He identifies six criticism of a truth-based epistemology. While assessing them rather cursory, he concludes that none of them provides a serious threat to his veritistic social epistemology. Those six criticism and some of their major proponents are: the argument from social construction (Richard Rorty, Bruno Latour), language and worldmaking (Jacques Derrida, Nelson Goodman), the unknowability criticism, the denial of epistemic privilege (Rorty), the argument from domination (Michel Foucault) and the argument from bias (Sandra Harding). Goldman's brief dismissal of this variety of different approaches has been criticized even within the field of social epistemology, most notably by Longino (Longino 2002c) and Kusch (Kusch 2002). Nonetheless, Goldman feels save to conclude that "[...] all the central arguments against veritism that spring from postmodern and constructivist quarters have failed" (Goldman 2003: 40) and that the door is therefore wide open for a veritistic social epistemology.

After having argued that the notion of truth indeed is useful and should be crucial in epistemology, one still has to decide *which theory of truth* is most adequate. Goldman defends a *correspondence theory of truth* that includes some elements of deflationary approaches (Goldman 2003: 41). The crucial aspect of such a correspondence theory is that "what *makes* sentences or propositions true are real-world truth makers" (Goldman 2003: 68). He defends this choice against pragmatic, epistemic and relativistic accounts of truth and proceeds to develop his veritistic social epistemology.

Veritistic Social Epistemology: Assessing Veristic Values of Social Practices

As noted before, epistemology quite generally for Goldman is "[...] a discipline that evaluates practices along truth-linked (veritistic) dimensions", and "[s]ocial epistemology evaluates specifically social practices along these dimensions" (Goldman

2003: 69). And his *veritistic social epistemology* is specifically targeted at supporting truth and countering error and ignorance. The central concept is that of *veritistic value*.

Goldman argues that in everyday-life as well as in science, certain value is placed on having true rather than false beliefs and this value is what he calls *veritistic value*. For his assessment of epistemic social practices it is essential that these a) can have different veritistic outcomes, such as knowledge (positive veritistic value), error (negative veritistic value) or ignorance and b) that they can be evaluated according to these outcomes. Accordingly, an epistemic practice has a higher veritistic value if it yields more knowledge than another epistemic practice. The goal of epistemology in general is to evaluate practices along truth-linked dimension and the goal of social epistemology consequently is to identify those social practices that have a comparably favorable effect on knowledge as contrasted with error and ignorance. *Epistemic states*, knowledge, error and ignorance have *fundamental veritistic values*. If someone believes a true proposition, this has the veritistic value 1 (=knowledge). If someone rejects a true statement, this has the veritistic value 0 (= error). And if someone suspends judgment on a true statement this has the veritistic value 0.5 (=ignorance) (Goldman 2003: 89).

Epistemic practices, by contrast, have *instrumental veritistic values* for achieving these states. Through different epistemic practices the overall veritistic values in a society, i.e. the relationship between truth, error and ignorance, can be changed. And these changes in turn can then be used to assess the value of epistemic practices. To illustrate how this evaluation of the veritistic value of different strategies might be achieved, Goldman (Goldman 2003) uses the example of *judgment aggregating* and compares different mechanisms of combining divergent expert opinions. He depicts a local weather bureau having five experts for the weather forecast (Goldman 2003: 81f). Those five experts unfortunately judge the likelihood that it will rain the next day very differently. Moreover, they have different forecasting *competencies*, conceptualized as different probabilities of delivering correct predictions. How should their opinions be amalgamated to provide the best possible weather prediction? Goldman offers three different strategies: The *unweighted majority rule*, the *dictatorial rule,* i.e. the bureau chooses the prediction of the most competent expert, and the *weighted voting*. In the case of weighted voting, the predictions of the experts are weighted by their competencies and as such are all amalgamated into a single value. Goldman demonstrates that, given the competencies are known,[78] the weighted voting scheme has the highest probability to lead to a true result. Modeled after this example, other social practices should also be evaluated with respect to their instrumental veritistic value, i.e. their ability to generate true predictions. Consequently, in the second part of his book, Goldman analyses different social practices (testimony, argumentation, technology and economics of communication, speech regulation and market place of ideas) with respect to their effects on the overall veritistic situation.

Generic Social Practices

Testimony

Testimony in Goldman's account refers to the process of communicating knowledge from one person to another. Stating that testimony is the most elementary and universal path to knowledge, Goldman focuses on the different roles and options for action that the provider and the recipient of testimony have. He conceptualizes testimony as a 4-stage process. After the initial discovery of knowledge (stage 1), someone decides to transmit a message containing this knowledge to others (stage 2). This message has then to be received (stage 3) and accepted (stage 4) by others. Only if all four stages are mastered successfully, knowledge has been transmitted from one person to another. This model strongly resembles Shannon & Weaver's mathematical theory of information (Shannon 1948). And in line with such a model of testimony as *signal transfer*, Goldman considers testimony to be responsible solely for the spread of knowledge, which has been created before. The crucial question from a veritistic point of view then is under what conditions the report of testimony will produce the largest possible change of veritistic value Goldman 2003.

If knowledge has been created before, the starting point for testimonial transmission of knowledge lies in the testifier's decision to convey this knowledge to someone else. Goldman bases this decision about whether and what to report to others on the newsworthiness of the information and the degree of change in veritistic value to be expected from this information transmission. The crucial term here is 'newsworthiness',

[78] This assessment and quantification of competencies is a crucial, but at the same time highly problematic aspect of Goldman's proposal. I return to this issue and the general difference between weighted and unweighted aggregation in the next chapters.

which in turn is defined as not being predictable on the basis of available background information.

Given that the information is transmitted correctly and received by a recipient, the epistemic task for the recipient of testimony is then to decide what to do with this information. He has to decide whether and to what degree to accept or reject the testimonial information or withhold judgment. Since the central task for a veritistic evaluation is to identify epistemic practices that lead to the highest veritistic gain, Goldman seeks to identify those testimonial practices of testifier and recipient that would best improve the overall veritistic situation and lead to largest increase in truth known in a community. So ideally, he seeks for a *general strategy* for accepting or rejecting testimony which is valid in all situations.

Unfortunately, in different reporting environments, different acceptance strategies concerning testimony may lead to the best results. For instance in the unlikely case that everybody always speaks the truth, the best acceptance strategy would be *blind trust*. In contrast in the reverse environment where everybody constantly lies, *anti-trust* would be the best strategy.[79] Since most reporting environments lie somewhere in between these two extremes different strategies are the advisable depending on the circumstances. However, given Goldman's interest in a general theory of testimony reception, judging case by case may not be a satisfactory solution. Instead Goldman asks whether there is an acceptance strategy imaginable, which *on average would work best in all possible reporting environments*. He argues that the application of *Bayes' Theorem* provides exactly such a strategy. Bayes' Theorem describes a probabilistic inference strategy which indicates how the probability of an event should be assessed and changed in the light of new evidence. To calculate this a posteriori probability of an event x given new evidence such as the testimony of a witness, what is needed are the prior probabilities of the x and not-x, as well as the conditional probability of the witness testifying x given that x has occurred and given that x did not occur. Please note that these latter conditional probabilities imply an assessment of the competency and truthfulness of a witness. The testimony of a testifier would be weighted by his perceived competency and truthfulness as was the case for the experts in the weather forecasts example.

[79] Of course this strategy would only work if the lie would always consist in the direct opposite of the true statement.

To conclude, according to Goldman (Goldman 2003), the application of Bayesian rules of inference promises to yield more true beliefs on average than most other procedures and thus should be employed in rational reasoning in order to maximize the number of true beliefs. Thus, Goldman considers the application of Bayesian inference to testimonial evidence to be a socio-epistemological, normatively appropriate practice. He turns this rather individual act of statistical inference into a socio-epistemological practice by simply applying it to information received socially via testimony. And by doing this, he argues for the epistemic value of weighting different agents according to their perceived competence and honesty.

Argumentation

In his chapter on *argumentation* Goldman then puts the information provider into the spotlight and asks how she should behave to be of high veritistic value herself (Goldman 2003). At first, he differentiates between *monological argumentation, dialogical argumentation* and *debate.* In a *monological argumentation* one person delivers an argument by giving premises and conclusions, in a *dialogical* one two persons argue about an issue and in a *debate* an additional audience is involved. With respect to different purposes of argumentation Goldman further distinguishes between *factual* arguments about what to *believe* and *practical* arguments about what to *do*. He confines his analyses to the factual cases.

In line with his focus on veritistic gain, he then analyzes in which cases which forms of argumentation lead to veritistically good results and comes up with *fourteen rules of good argumentation*. He argues that these rules are not his invention, but tacit folk rules that ensure cooperation and information sharing in communities. For monological arguments to be veritistically beneficial, the following criteria should be met: Be informative, be credible, be comprehensible to your audience. Additionally it would be good if no one in the audience can defeat your argument and the remaining rules center around questions of how to deal with defeaters and rebuttal as well as ho to respond to criticism. Goldman puts emphasis to what he calls the truth-in-evidence principle (TEP), which holds that "[a] larger body of evidence is generally a better indicator of the truth-value of a hypothesis than a smaller, contained body of evidence, as long as all the evidence propositions are true and what they indicate is correctly interpreted. (Goldman 2003: 145)".

He argues that this principle implies that it is veritistically desirable to engage in critical debates. If being critical is an epistemic duty, a crucial question for any social epistemology therefore consists in finding ways to support and foster an intellectual climate in which criticism can flourish. Goldman's conclusions are straightforward and one of the few similarities with Helen Longino's approach[80] (Longino 2002c): increase the incentives for criticism and expand the opportunities for it.

He concludes his normative-prescriptive account or argumentation by assessing four argumentative fallacies in the light of his veritistic social epistemology: *argumentum ad verecundiam*, i.e. the appeal to authority*, begging the question, argumentum ad hominem*, i.e. attacking the person instead of the argument, and *the straw man fallacy*, i.e. a misrepresentation of an opponent's position to make it easier attackable. While from his veritistic point of view the *appeal to authority* is not a fallacy in cases in which the authority really is knowledgeable, *begging the question* is problematic because no new true conclusions are conveyed. And while *argumentum ad hominem* might be admissible if there are serious doubts about the competency and integrity of an opponent, the *straw man fallacy* controverts the *principle of accuracy* and should therefore be avoided on veritistic grounds.

Information and Communication Technology (ICT)

The third generic social practice Goldman analyses in the light of his veritistic social epistemology are information and communication technologies (ICT). The major criterion of his Goldman's assessment of the merits of ICT again lies in the potential of these technologies to increase the overall veritistic value. This can be done in two ways: either, more people get to know something, or something new gets known, i.e. new knowledge gets discovered. As was already the case in his analysis of testimony, Goldman's focus lies in the spread of existing knowledge and not in the creation of new knowledge. The epistemological task then consists in finding those means that can spread knowledge *fast to many people*. Besides the speed and breadth of spreading knowledge it might be of additional epistemic value that knowledge can be easily found and differentiated from non-knowledge. This aspect of *quality assessment* is increasingly relevant today where it is rather information overload than information scarcity which is problematic. Goldman analyses different tools, such as email, forums,

[80] Helen Longino's approach is portrayed in detail below.

the WWW more generally, search engines, etc with respect to their contribution to veritistic gain and shows how those media can either be used to raise the overall veritistic value in a society by distributing knowledge or to lower it, e.g. by spreading error.[81]

ICT and Scholarly Communication

An interesting issue that Goldman specifically addresses concerns the role of ICT for *scholarly communication*. The main topic he addresses is *peer review* and the difference between peer reviewed journals and e-print servers such as (http://arxiv.org/). Goldman argues that the process of *peer review* is beneficial for readers because experts filter and evaluate information for them. And for writers it is beneficial, because they gain reputation by having passed peer review and by publishing in those journals that have strict evaluation criteria. The fact that e-print servers without peer review function so well in certain natural sciences is explained by him through the high entrance barrier into sciences such as physics. Moreover, since the claims in physics are in principle replicable and the methods are verifiable, physicists would not dare to publish something that does not meet their community's standards on those servers. Thus although, there is no peer review, quality can be controlled in this case through the *fear of post review*. A crucial prerequisite for this mechanism to work lies for Goldman in identification. One has to be able to identify the author of a paper on such non-peer-reviewed severs, because otherwise the fear to lose one's reputation by publishing low-quality content does not work. Anonymity, as is the case for instance on Wikipedia would be detrimental on such an account. Further issues that Goldman raises with respect to scholarly communication are the costs of online publishing, the dangers of pay-per-view-policies for libraries as well as the balancing of different values, such as property rights versus public interest or freedom of information. These topics are currently hotly debated in academic discourse and I return to some of them in Part 3.

Speech Regulation & the Marketplace of Ideas

Goldman concludes his analyses of generic social practices with a chapter on *speech regulation and the marketplace of ideas* (Goldman 2003). Basically in this chapter he is

[81] Since many applications Goldman describes are by now outdated, I do not get into detail in describing them. However, Goldman has published a more recent paper on blogging (Goldman 2009) to which I return in the next chapter.

analyzing general policies for allocating speech opportunities with respect to their impact on the overall veritistic value in a society. Speech opportunity is to be understood in a broader sense of giving people the change to be heard or read, i.e. it does not only include direct allocation of speech time in a face-to-face encounter, but also the possibility to publish, broadcast, etc. Crucial issues he raises concern the relationship between facilitation speech and the prevention of speech (e.g. the rejection of a paper by an editor). In particular he analyzes the free-market thesis, according to which the market has to regulate itself and state intervention has to be avoided, and its impact on the veritistic value. In conclusion, he rejects the idea that an unregulated "knowledge market is veritistically superior and argues instead that some form of regulation (e.g. state intervention into what is broadcasted, peer review, editorial choice, etc.) is needed to enable the *prerequisites* of a free market of ideas in the first place. Which forms of intervention are paramount is left open. But he insists for instance, that the inclusion of many people into epistemic practices is not *per se* desirable. Quite to the contrary, speech control might be necessary at times and with respect to science, Goldman concludes that the filtering function of peer review is a clear example of the benefits of such regulation and denial of speech Goldman 2003.

Science

From the four special domains that Goldman applies his veritistic social epistemology to, I focus on science only in more detail. One reason for this choice is that science is most directly concerned with knowledge creation. As a consequence, it is the topic which most other social epistemologies portrayed in this chapter have focused on and the differences between the different approaches can be made most transparent when a common framework is chosen.

Goldman sets on what turns out to be a *defense of the veritistic superiority of science* with a critique of two seminal publications in the history and social studies of science: Steve Shapin and Simon Schaffer's "Leviathan and the Airpump" (Shapin and Schaffer 1985) as well as Bruno Latour's "Science in Action"(Latour 1987). The lense through which he reads those approaches is again his veritism. In "Leviathan and the Airpump", Shapin and Schaffer examine the debate between Robert Boyle and Thomas Hobbes over Boyle's air-pump experiments in the 1660s with respect to the role of societal factors related to the different knowledge systems promoted by Boyle and Hobbes

(Shapin and Schaffer 1985). Goldman interprets Shapin and Shaffers analyses of Boyle's experimental program as a *critique of credentialism*, i.e. the reliance upon credentials such as titles as an indicator of a person or group's intellectual worth and as a proxy to assess their knowledge claims. Goldman argues that from a veritistic point of view credentialism might not be problematic at all. Quite to the contrary, he asserts that a "[...] properly applied practice of credentialism would promote rather than hinder veritistic ends" (Goldman 2003: 225).

The second approach, which Goldman seeks to debunk is Latour's empirical account of science in the making as depicted in *Science in Action* (Latour 1987). One the one hand, Goldman (mis-)interprets Latour as arguing that science is nothing but politics. On the other hand, he argues that Latour's "political-military account of science" is in principle compatible with a veritistict epistemology, because deference to authorities can be epistemically valid if those authorities are reliable (Goldman 2003: 225ff).

Goldman continues in his defense of science by arguing that the role of *bias* in science is often exaggerated (Goldman 2003: 234). To support this claim, he turns to experiments from the cognitive sciences in which *cognitive biases*, such as *overconfidence*, could have been reduced or overcome by minor changes in the instructions (e.g. Gigerenzer 1991, cited from Goldman 2003). As for the non-cognitive *hot biases*, he confines himself to argue that *doxastic voluntarism*, the notion that "one can *choose* to believe or disbelieve by a simple act of will" is controversial (Goldman 2003: 234) and that *wishful thinking* can be countered in science even better than elsewhere with the right incentives. As for values and cultural perspectives and their impact on science, Goldman simply asserts that such *societal biases* although they exist, exert less influence in science than in other societal systems. After a brief detour through the problems of underdetermination and the theory-ladenness of observation, Goldman lays open to be a scientific realist, whose social epistemology can nevertheless also be adopted from an anti-realist position.

He continues his veritistic analyses of science with an argument for the *comparative scientific superiority*, which he defines as follows: "Scientific practices are veritistically better than any set of nonscientific practices available to human beings for answering the sorts of questions that science seeks to answer"(Goldman 2003: 247). This superiority is related to six characteristics of science, which distinguish it from other practices and domains. I list these characteristics of science, because even if one does not consider them to be the "sources of scientific success" (Goldman 2003: 250), they nonetheless stress some important features of contemporary science. Therefore, they are not only important for Goldman's own socio-epistemic account of science, but they also relate it to other social epistemologies to be portrayed in this chapter.

1. "An emphasis on precise measurement, controlled test, and observation, including a philosophy, organon, and technology for more and more powerful observation.
2. A systematic and sophisticated set of inferential principles for drawing conclusions about hypotheses from observations of experimental results.
3. The marshalling and distribution of resources to facilitate scientific investigation and observation.
4. A system of credit and reward that provides incentives for workers to engage in scientific research and to distribute their efforts in chosen directions.
5. A system for disseminating scientific findings and theories as well as critical assessments of such findings and theories.
6. The use of domain-specific expertise in making decisions about dissemination, resource allocation, and rewards."(Goldman 2003: 250f)

Goldman argues that from those six characteristics, at least the last four are social. Therefore, a socio-epistemology of science has to assess *whether those social aspects of science are actually beneficial or detrimental for science.* Moreover, for each of those characteristics, there might be better or worse practices and a social epistemology has to assess the *respective veritistic value of different social scientific practices*.

How could such a veritistic analysis of social practices in science look like? The first social aspect of science that Goldman addresses concerns the *distribution of scientific labor*. Goldman critically assesses Philip Kitcher's perspective on the distribution of cognitive labor according to which the diverse motives of scientists are sufficient to ensure a distribution of cognitive labor that is beneficial for science (Kitcher 1993).[82] More specifically, Kitcher introduces the term notion of the *epistemically sullied agent*, a credit-driven scientist whose primary motivator is *being the first* one to solve a scientific puzzle rather than just aiming at truth in general. While Goldman agrees with

[82] The debates on the distribution of cognitive labor in general and Kitcher's work in particular are portrayed in more detail in the next Chapter.

Kitcher, that also from a Bayesian perspective, a certain diversity of approaches is beneficial and should be supported, he differs in his assessment of the value of the sullied agents and their drive for credit. Goldman asserts that credit-orientation, although it may not *necessarily* be detrimental to science, can make scientists chose a veritistically inferior research design (Goldman 2003: 263).

Goldman then continues to give recommendations concerning *scientific publishing*, e.g. on the question of how editors should chose articles to increase the overall veritistic value in a society. Based on his prior analyses he concludes that editors should choose those articles which are not only qualitatively best, but also those which are the most *significant*, because those articles lead to the *highest veritistic gain.* And to assess the potential impact of scientific contributions, editors should turn to authorities in the field. Especially in cases of highly surprising results, i.e. those results with the highest potential to change the veritistic distribution, expert assessment is crucial, because experts are the most likely ones capable of assessing whether the results are really groundbreaking or rather due to mistakes in the research design, a statistical exception, etc.

Given this crucial role of experts, it should not come as a surprise that Goldman puts emphasis on analyzing the relevance of *recognizing authorities* in science, a topic on which he elaborates in a later article as well (Goldman 2001). In arguing that everybody can assess expertise, i.e. the quality of different experts, he seeks to refute John Hardwig's doubts concerning the ability of lay people to assess expertise (Hardwig 1985). Hardwig's work on trust gets portrayed in the next chapter in more detail. For the moment it suffices that Goldman cites Hardwig in arguing that lay people cannot assess experts empirically, because they "[...] do not fully understand what constitute good reasons in the domain of expert opinion" (Hardwig 1985). Lay persons cannot assess the quality of an expert precisely because as non-experts they do not possess the competencies to assess the expertise of others. Hardwig concludes that therefore, laypeople sometimes simply have to *trust* experts, that this *trust is sometimes even an epistemic duty.* In a subsequent article he further argues that *scientists themselves have to trust their peers* in order to successfully operate in science (Hardwig 1991). Goldman (Goldman 2003) claims that this fallback on trust is precipitate and that lay people are perfectly capable of deciding about the merits of different experts if they follow some rules. He argues that even without being an expert, people can assess whether an expert's predictions came true, they can verify the expert's claims elsewhere, they can assess whether the argumentation was sound and they can compare the conclusions of different experts. I return to Goldman's perspective on expertise and credentialism as well as to the topics of epistemic authority, trust and reputation more generally in the next chapter.

5.4.3 Miriam Solomon's Social Empiricism

Introduction

Miriam Solomon is another philosopher who has offered a comprehensive social epistemology which she has labelled *Social Empiricism*. Similarly to other social epistemologists, she acknowledges that the sociology of scientific knowledge, feminist studies as well as various historical, ethnographical or psychological accounts of science have posed serious challenges for traditional epistemology and philosophy of science. She takes that in order to develop more adequate and fruitful theories to account for the sociality of knowledge, the insights and claims coming from those fields have to be taken serious, and states that her approach can account for these challenges (Solomon 2001).

Solomon argues that despite other differences most philosophers and sociologists of science share not only standards for the evaluation of scientific practises, but also many underlying assumptions about science due to a shared background in Enlightenment epistemology (Solomon 2001: 2). These shared premises include *individualism*, i.e. the focus on the individual thinker; the demand that *science should be free of motivational or ideological bias*; the *appreciation of consensus* and the *all-or-nothing-quality of rationality*, i.e. the credo that rationality does not come in degrees. Solomon argues that these "shared assumptions about the nature of rationality and progress lead to mirror image views of the nature of scientific change" (Solomon 2001: 6) and that they need to be overcome for a new *more social epistemology.*[83] However, as is shown in the course of this analysis, instead of proclaiming her theory to be *more social*, it may be more adequate to label her approach (social) *macro-epistemology*. Similar to Steve Fuller's account Solomon's social epistemology aims at offering guidance for science policy.

[83] An earlier article by Miriam Solomon introducing her *Social Empiricism* is entitled "A more social epistemology" (Solomon 1994).

The perspective she takes is that of an external observer who evaluates epistemic activities and their distribution within a scientific community. Via funding, the distribution of such activities could then be changed according to socio-epistemological considerations. Her assertion that *Social Empiricism* is a social epistemology for science only and that it is more normative than descriptive (Solomon 2001) fits this alignment with the goals of science policy.

Key Features

Solomon delivers a full account of her social epistemology in her book "Social Empiricism" and unless otherwise noted my portrayal is based on this book (Solomon 2001). For the development of her theory, she uses case studies in the history of science to analyze the effects *different types and distributions of bias* had on the development of science in different fields. More specifically, she focuses on the role of biases for the achievement and dissolution of consensus on different theories within scientific communities, their effects on the distribution of cognitive labour and the overall impact on the advancement of science in different scientific fields measured by empirical success. In focusing on the *distribution* of these biases, it is the scientific community instead of the individual scientist that is in the centre of her inquiry. It is in this shift of attention to the scientific community as a whole that she claims to overcome the epistemic individualism that has long hampered philosophy of science and epistemology.

Solomon's above mentioned critique of the shared premises of philosophers and sociologists of science also serves as a rationale for the development of her own account that sets on to overcome this individualism in a quite distinct way. In the following I introduce the three central socio-epistemic topics that are at the centre of Solomon's approach. The first one concerns the *distribution of cognitive labour* at the level of scientific communities. The second related topic concerns the role of *consensus and dissent* within scientific communities. And the third topic concerns the effects of *different types and distributions of biases* on the development of science. Instead of using the common, but rather pejorative term "bias", Solomon introduces the notion of "decision vectors". This move is related to her argument that biases are not per se detrimental to science. Instead, it depends on the *distribution* of different types of biases whether science is advanced or hindered.

Division of Cognitive Labour: The Social Coordination of Epistemic Effort

Division of cognitive labour concerns the question of how epistemic work should be most rationally and epistemically beneficial distributed over a multitude of people. This issue can be tackled on the micro-level of deciding upon how many people and who works on which tasks within a research group just as much as on the macro-level of deciding about the funding for different fields of research. Solomon's focus lies on the level of a research community and the guiding question is how many, which and for how long different research strategies should be pursued. This perspective that she adopts is comparable to a queen sending different variously well equipped ships into different directions to discover new continents. In a more modern variant, it is the view of science policy makers who have to decide upon the distribution of research grants, about the amount of money to be spent on different fields of research and different research strategies within these fields.

This focus on the distribution of cognitive labour within scientific communities – not within research teams - is the framework for Solomon's *Social Empiricism.* Its goal is to develop guidelines for science funding based on the epistemological analysis of various case studies in the history of science. In effect, Solomon aims at culling examples from the history of science, in which the overall epistemic situation is characterized by *advantageous distributions of decision vectors*. Then these retrospective insights from historical analyses have to be transformed into prospective recommendations for science policy that ensure that such favourable epistemic situations with advantageous bias patterns will be achieved. According to Solomon, cognitive effort should ideally be distributed *equitably*, i.e. proportional to the em*pirical success* of different theories or methods. Since empirical success however, can only be assessed *after* research has been conducted, her recommendation concerning the funding of research alternatives remain quite simple: As long as not all empirical evidence can be accounted for by one theory, all other theories that have empirical success which this theory cannot account for should continue to be pursued and thus have to be funded. This means that premature consensus has to be avoided as long as there is at least some empirical success by theories other than the mainstream theory.

Consensus and Dissent

The advice to fund diverse approaches and to avoid premature consensus on a specific theory or method leads to the second major topic of Solomon's *Social Empiricism*: the questions of consensus formation, retention and dissolution in science, the role various social factors play for it and the effects it has on the development of science in different fields. To get a grip on these questions, Solomon analyzes case studies in the history of science with respect to when, under what conditions and with which effects - consensus was achieved, retained or dissolved.

For Solomon dissent plays a much bigger and more positive role in science that normally assumed. She considers consensus to be a special case of dissent, namely zero degree of dissent and argues that while dissent was traditionally perceived to be only a temporary phenomenon, that disappears once the source or error is detected, dissent usually is much more profound and enduring. In those traditional accounts, it is more or less assumed that if all scientists have the same information and act rationally, cognitive uniformity is to be expected - at least on the long run. Reasons for existing dissent continue to be numerous and diverse: imperfect communication of information, different access to evidence or different subjective prior probabilities for different theories, different weights for each of their theoretical values (e.g. fruitfulness, predictive accuracy, etc.) or simply different methodologies. However, once enough evidence is gathered and information has been properly communicated, etc. it is expected that there will eventually be consensus.

That consensus is something inherently valuable also lies behind several normative theories of consensus, such as Keith Lehrer's theory on rational consensus formation (Lehrer and Wagner 1981), which is portrayed in the next chapter. Consensus in those normative approaches functions as an attainment, a major goal and endpoint of scientific inquiry. Presumably once truth has been identified and consented upon in a scientific community, social factors supposedly do not play any role anymore and hence, there is no need for a social epistemology at this stage anymore. This is where Solomon disagrees. Instead of taking consensus to be intrinsically valuable and the ultimate goal of science, Solomon shows that there have been numerous cases of *premature* consensus in the history of science, cases in which an early agreement on research agendas, on methods and theories, has actually hampered science by precluding the pursuance of alternative approaches. She argues that while there certainly are some cases in which consensus is and remains normatively appropriate over time, in the majority of cases there either should have been still considerable dissent at the time consensus on one theory or method was reached or consensus should have been dissolved quicker in the light of contradicting evidence.

One of Solomon's examples is the 'central dogma' in molecular biology, which holds that it is only the DNA in the nucleus, which is in control of cellular processes via mRNA and protein synthesis. This exclusive focus on the nucleus has led to a neglect of cytoplasmic inheritance and the processes of non-chromosomal transmission taking place in the rest of the cell. The result was not only to ignorance, but also to the depreciation of those researchers who worked on such non-mainstream questions. According to her analysis it took molecular biology decades to loosen the grip of the central dogma to get non-nuclear inheritance into the focus of mainstream research. Thus, in this case, premature consensus, i.e. consensus on one theory that could not account for all empirical evidence, had a negative effect on the advancement of science, because it rendered the pursuance of alternative approaches unpromising. By doing this, it had an unfavorable effect on the distribution of cognitive labor within this research community and hindered research that would have been able to amend and improve mainstream research in molecular biology.

One of the major goals of her *Social Empiricism* accordingly is to deliver some guidance as to whether and when to refrain from consenting, when to dissolve consensus and how to decide on the distribution of cognitive labor over different approaches. The bottom-line of her argument is that as long as more than one theory or method is empirically successful, consensus on the mainstream approach should be suspended, because such premature consensus formation would be detrimental to the advancement of science. Instead, research efforts should be distributed around these different empirically successful theories – ideally in an equitable way.Thus, in contrast to classical normative theories of consensus such as the one developed by Keith Lehrer, she is not recommending a new method of consensus formation, but rather argues for a suspension of consensus and the pursuance of different research strategies until finally *all empirical evidence is supporting only one theory*. Since this condition is hardly ever met, the safest strategy is to support different research agendas for as long as possible, i.e. to support scientific pluralism.

Empirical Success

As indicated by the title of the book, Solomon is strongly committed to empiricism and as such, her approach privileges *empirical success* over *theoretical success.* By empirical success, Solomon means predictive, retrodictive and also technological success. Theoretical success denotes concepts such as simplicity or elegance of a theory, breadth of scope, but also Helen Longino's list of feminist theoretical virtues, e.g novelty, ontological heterogeneity (Longino 1995). She argues for the primacy of empirical success for the following reason: empirical success is due to dependable behaviour of the world and as such not entirely fabricated or man-made. Moreover, it is indispensable: if empirical success is not a value, it is not science. Theoretical success by contrast, is not only contingent to the inquirer herself and accordingly less valuable, it is also negotiable. She concludes that this is why "[e]mpirical success is a primary goal of scientific inquiry, and theoretical success valuable only when it brings extra empirical success, convenience or moral benefits with the available empirical success" (Solomon 2001: 20).

Considering the centrality of the term empirical success, but also Solomon's more general tendency to objectify science policy, it is astonishing that the term empirical success remains quite vague. Evaluating other accounts on empirical success, she concludes that definitions of empirical success are of little use and confines herself to describing it as the moment when "[…] scientists, instruments and the world successfully coordinate their actions as a result of tinkering, conceptual adjustment and serendipity" (Solomon 2001: 27f).

Whig Realism

Even though Solomon stresses the importance of empirical success, she nonetheless agrees with Goldman (Goldman 2003) and Kitcher (Kitcher 1993) that *truth*, also serves as one of the primary goals of science. However, taking into account historical analyses, she rightly concludes that truth might be a difficult concept for assessing scientific theories, because the majority of claims and assumptions in science turn out to be false over time. Nonetheless, instead of rejecting the relevance of truth for science or developing an alternative epistemic success term, she sticks with the notion of truth and labels her approach *Whig realism*, as a blend of Whig history [84]and realism. *Whig realism* is meant to denote that although scientific theories might neither be literally, nor partially or approximately true nor even good representations, there is typically *something true* about empirically successful theories (Solomon 2001: 11). Giving a positive definition of her concept of truth however, seems to be more difficult, so that the following is as close as we get to a definition about what she means by *truth in the theory:* “What is true about our empirically successful theories is, typically, an implication of the theory at the theoretical level (i.e., not just a prediction or observation) that may or may not be explicitly derived during the historical period in which the theory is accepted.” (Solomon 2001: 39). Crucially, this *truth in the theory* can only be recognized in hindsight. Thus, this truth can be analyzed and ascribed only by historical reconstruction, but is impossible to assess in the course of the actual research process. Only after having assessed the truth value of theories retroactively and having compared the different distributions of bias can we decide upon the benefits or detriments of different distributions and draw normative-prescriptive conclusions from them for science policy. This insight also explains Solomon’s methodological choice of using case studies to make normative recommendations for science policy.

Solomon concludes that her *Whig realism* is superior to realist and anti-realist accounts of truth, because of this possibility to make concrete recommendations for science policy based on historical analyses. More precisely and with reference to the question of consensus, she argues that while traditional realists simply aim at consensus in science, and anti-realists are indifferent to consensus or dissent, Whig realism can be more specific in arguing for or against consensus in different situations. Quite generally, consensus is only very rarely normatively appropriate, namely only when one theory can account for *all empirical evidence*. Otherwise, Solomon’s approach would demand an encouragement of pluralism up until the point at which all last doubts about a theory are dispelled and all empirical evidence yields into one direction. Before returning to the questionwhen exactly *dissent* is normatively appropriate, we need to take a look at her concept of decision vectors.

[84] The term *Whig* goes back to the Whig party which has been one of the main parties in the British Parliament. *Whig history* usually denotes interpreting history from the present point of view and as a story of progress toward the present. This often also means implying a line of causation and stressing the necessity and unavoidability of historical development (Solomon 2001: 33).

Decision Vectors: The Role of Bias

The key concept that Solomon introduces to account for the effect of various social factors on science and knowledge are her so-called *decision vectors*. In trying to show that social factors are not inherently detrimental to science and epistemic practices more generally, she proposes this term as an alternative to the clearly pejorative notion of *bias*. Following Hume and consonant with Kitcher's conviction that "[p]articular kinds of social arrangements make good epistemic use of the grubbiest motives" (Kitcher 1993: 305, quoted from Solomon 2001: 53), she argues that the term decision vector allows for biases having a positive effect on science.

Decision vectors denote *all factors* that influence scientific decision making and as such have an effect on scientific outcomes. For Solomon, they include such diverse things as ideology, pride, peer pressure, deference to authority, salience of data or even the birth rank of scientists. Allegedly based on her case studies, but without giving any support for this claim, Solomon contends that there are 50-100 decision vectors and that while each of them is neither beneficial nor detrimental to science, their impact depends on their overall distribution (Solomon 2001).

There is only one major differentiation she proposes, namely the difference between *empirical and non-empirical decision vectors*. The importance difference lies in their relation to empirical success. While empirical decision vectors are "[…] causes of preference for theories with empirical success, (Solomon 2001: 56)", non-empirical decision vectors lack this connection to empirical success. Examples of such empirical decision vectors are *salience and availability of data* as *cognitive* factors, an *egocentric bias* towards one's own data as a *motivational* factor as well as more generally a preference for a theory which generates *novel predictions.* Non-empirical decision vectors by contrast are not related to empirical success and include social and political factors (ideology, deference to authority, agreement with scripture), motivational factors (pride, conservativeness, radicalism, competition, peer pressure) and cognitive factors (representativeness heuristic) as well as theoretical values, such as elegance, simplicity, or Longino's feminist epistemic virtues (Longino 1995).

Relating Decision Vectors, Consensus and the Distribution of Epistemic Labour

How then are these decision vectors related to the formation of consensus and the distribution of cognitive labour? The connection between consensus and the distribution of scientific effort is straightforward, given that premature consensus can lead to an untimely cutting of research grants for alternative approaches and an accordingly unbalanced distribution of cognitive effort. The relationship to decision vectors however, is not that clear. By providing various historical examples from evolutionary biology, genetics, plate tectonics and cancer research, Solomon shows that decision vectors have played a role in the formation of consensus (Solomon 2001). More specifically, different distributions of decision vectors have been either beneficial or detrimental to the advancement of science in these different fields by affecting consensus formation or dissolution.

Beneficial distributions of decision vectors would lead to consensus only when all empirical evidence can be accounted for by one theory. Otherwise they would ensure scientific pluralism. The positive example from the history of science that she cites here is plate tectonics, where three different theories have been pursued in parallel until finally continental drift was able to account for all empirical success. Detrimental distributions of decision vectors, by contrast, would either lead to premature consensus (as was the case concerning the central dogma in molecular biology), to delayed dissolution of consensus or possibly also to a refusal or delay of consensus, after one theory actually has already accounted for all empirical evidence.

Advantageous Distributions of Decision Vectors: Guidelines for Science Policy

But how exactly do such beneficial distributions of decision vectors look like or how *should* they look like in comparison to rather detrimental ones? And even more importantly, how can they be made useful for science policy? Solomon states that the ideal distribution consists in an *equitable distribution of empirical decision vectors combined with an equal distribution of non-empirical decision vectors*. She then simply *counts* those decision vectors she identifies for each theory in each historical case study, adds them up and compares their ratio for different competing theories. She concludes that the closer the distribution of decision vectors was to this ideal distribution, the better science developed. If the decision vectors were far from ideal, problems such as premature consensus and unbalanced distributions of cognitive labour were the result. The goal of science policy would then be to affect the distribution of decision vectors to approximate this ideal state.

By arguing for a science policy that changes the distribution of decision vectors, Solomon rejects an invisible-hand model of scientific reasoning.[85] Similar to Fuller's argument (Fuller 1988) and against Kitcher (Kitcher 1993), she asserts that science does not self-organize in the best possible way without corrective intervention. Rather, science policy makers should intervene by making use of insights from a normative social epistemology that can prescribe how to ensure such favourable distributions of decision vectors (i.e. equitable distribution of empirical decision vectors, equal equitable distribution of non-empirical decision vectors. Solomon argues that her *Social Empiricism* is exactly such a normative social epistemology. More specifically, she specifies under which conditions consensus, dissent and the dissolution of consensus are normatively appropriate and makes these recommendations the normative-prescriptive core of her *Social Empiricism*. These are her three recommendations concerning consensus, dissent and the dissolution of consensus. There are distinct conditions under which dissent is appropriate and others under which consensus would be adequate and these conditions are related to the distribution of empirical success and the distribution of the different types of decision vectors. According to Solomon, when the following three conditions are fulfilled, *dissent* should prevail:

> "1. Theories on which there is dissent should each have associated empirical success.
>
> 2. Empirical decision vectors should be equitably distributed (in proportion to empirical successes).

[85] When unfolding her concept of decision vectors, Solomon explicitly differentiates herself from other what she calls the "invisible hand model of scientific rationality", which she ascribes to Goldman (1992) and Kitcher (1993). She asserts that this model is named after Adam Smith, denoting "[…] that the impulse of self-interest, in a laissez faire economy, will bring about the desirable goal of public welfare (Solomon 2001: 55)". I think that she has fallen victim to the same misreading as many contemporary advocators of the 'wisdom of the crowds'. First of all, Smith's warning that state intervention might not be beneficial was set against the background of the very centralized France at the end of the 18th century, a situation which is hardly comparable to modern societal frameworks. Moreover, and even more importantly, Smith's is not simply advocating some form of free market liberalism. His recommendations emphasize the need for conscience and sympathy, the need to constrain one's behavior through virtue (moral sentiments), in order to balance the self-interest. Thus, Smith himself did not simply proclaim that everything settles for the best, if there only is no state interference. Rather, he based the success or failure of this on the virtue of sympathy.

> 3. Non-empirical decision vectors should be equally distributed (the same number for each theory)" (Solomon 2001: 117f).

Since she considers consensus to be a special case of dissent, namely zero dissent, her normative account of dissent can be applied to consensus as well. Consensus is normatively appropriate only if *one theory* can account for *all empirical evidence*, which hardly ever is attainable.

> "1'. One theory comes to have all the empirical success available in a domain of inquiry.
>
> 2'. This same theory comes to have all of the empirical decision vectors, since all scientists working productively (with empirical success) are working within the one theory.
>
> 3'. Any distribution of non-empirical decision vectors is OK, but typically more will develop, over time, on the consensus theory, as the old theories fade away. During dissent, and thus in the early stages of consensus formation, the above requirement of equal distribution of non-empirical decision vectors holds" (Solomon 2001: 119).

However, since Solomon argues that consensus should not be regarded the final stage of scientific development, she also needs to account for the dissolution of consensus. This process basically is set off by new empirical evidence being produced by a theory other than the consensual one. If as a result the empirical decision vectors then become equitably distributed and the non-empirical ones equally distributed, consensus should be dissolved. With her own words, in order to dissolve consensus the following prerequisites should be met:

> "1". A new theory has empirical success that is not produced by the consensus theory. (So, the new theory deserves attention.)
>
> 2". Empirical decision vectors come to be equitably distributed.
>
> 3". Non-empirical decision vectors come to be equally distributed" (Solomon 2001: 119f).

These three normative recommendations concerning consensus, dissent and the dissolution of consensus are at the heart of Solomon's *Social Empiricism*. She argues that her approach is *social* (and even more social than others) in making normative recommendations on the level on the community as opposed to normative recommendations for individual scientists. And it is *empiricist*, because *empirical success* along with *truth* is of prime importance. In comparison to other social epistemologies she argues that *Social Empiricism* demands more and less. While it does not demand that individual scientists improve their scientific reasoning, it requires changes at the macro-level, i.e. systematic changes of science produced by a science policy based on *Social Empiricism*. In this sense, *Social Empiricism* is a form of social engineering for science (Solomon 2001).

Thus, in the end, the normative implications of Solomon's *Social Empiricism* are quite clear: Suspend consensus on one theory for as long as any other theory also has empirical success that the mainstream theory cannot account for. For science policy, this means that alternative approaches should be funded up to that – quite unlikely - point of consensus as well. With respect to the ideal distribution of decision vectors, Miriam Solomon is also quite explicit: ideally empirical decision should be equitably distributed and non-empirical decision vectors should be equally distributed. However, when it comes to the crux of the matter this explicitness of profoundly lacking: How could such a favorable distribution of decision vectors get achieved? Which means of science policy are appropriate? How could science policy makers ensure that empirical decision vectors are equitably distributed and non-empirical decision vectors equally distributed? What is the use of quantifying decision vectors, given that the only option for action that science policy makers have consists in different amounts of money to different research fields and fund different approaches as long as dissent prevails?

Solomon concludes her book with acknowledging that "[s]ocial empiricism is conceptually simple. There is nothing mathematically or philosophically challenging in the idea that empirical decision vectors should be equitably distributed and non-empirical decision vectors equally distributed. Difficulties come in identifying decision vectors, and in making realistic recommendations for changing their distribution." (Solomon 2001: 151). Unfortunately, both these difficult tasks have not been tackled in this book. To play the devils' advocate one may question the relevance of *decision vectors* for Miriam Solomon's account altogether. If we skip the 2nd and 3rd criteria for consensus, dissent and consensus dissolution would not the normative recommendations stay exactly the same? Fund alternative approaches as long as they are empirically successful! That this funding should ideally be equitable might be a good idea, but we should remember that the empirical success can only be known after research is conducted. The concept of decision vectors does not help in finding out how to achieve such an equitable distribution. I critically assess Solomon's approach and its relevance for the analysis of socio-technical epistemic systems in the next chapter.

5.4.4 Martin Kusch's Communitarian Epistemology

Introduction

In his book "Knowledge by Agreement" Martin Kusch develops a *Communitarian Epistemology*, which he explicitly distinguishes from other social epistemologies portrayed in this chapter. Thus, while he asserts that his *Communitarian Epistemology* is not a social epistemology, he develops is dialogically against other social epistemologies. Those social epistemologies he refers to critically for him fall into one of the two categories of being either *complementary programmes* (e.g. Goldman 2003) or *science policy programmes* (most importantly Fuller 1988, Kusch 2002: 2ff). While the latter aims at improving science by changing its structures, the former seeks to develop epistemic accounts of social practices to *complement* individualistic epistemologies as traditionally conceived.

Kusch's *Communitarian Epistemology* is different and understands knowledge to be social in a much more fundamental sense than has been the case in the previous approaches. For its development Martin Kusch sets on to combine epistemology with political theory. Communitarianism in political philosophy denotes a priority of community over the individual in order of explanation. This means that "[m]oral individuals do not precede moral communities; moral individuals can be understood only through their membership in moral communities" (Kusch 2002: 1). For the development of a communitarian epistemology, this means that *knowers can also be understood only through their membership in epistemic communities*. Or to put it even more strongly, an isolated individual knower does not exist in a communitarian epistemology.

Kusch bases his *Communitarian Epistemology* on two premises concerning the sociality of knowledge, in which this central role of the community becomes obvious. First, terms such as "knowledge", "knower" and "know" mark *social statuses*, and by this they depend on communities to grant these social statuses. There are no exceptions to this rule. Without a community of knowers, there can be no knowledge. Secondly he claims that, "[…] the social status 'knowledge' is typically granted to, or imposed on, *groups* of people" (Kusch 2002: 1). In inserting the word "typically', Kusch allows for the possibility that someone outside a community might be attributed knowledge.

Nonetheless, although such an individual may possess knowledge, it can only be ascribed to him by a community.

With these two prerequisites, Kusch strongly deviates from the individualism which is still prevalent in the majority of social epistemologies, especially in those approaches, that he labels complementary. By arguing that the primary subject of knowledge is not the individual, but the community his communitarian approach goes as far as to conclude there is no such thing as the individual isolated knower. Individuals may have beliefs. But, beliefs can only become knowledge if a minimal *epistemic community* is formed around them. And by this process, these beliefs are not held by an individual any longer, but by a community bound together by a set of entitlements and commitments (Kusch 2002).

Kusch sharpens his account by developing it dialogically against other theoretical approaches that aim at accounting for the social context or the social nature of knowledge. These approaches for instance include Keith Lehrer's consensualism, Donald Davidson's interpretationalism or different variants of contextualism. The three main epistemological problems he addresses in this book are the nature of testimony, the rationality of empirical belief and questions of objectivity (Kusch 2002). For the purpose of this thesis, emphasis is laid on his disquisition of testimony. Within the development of his communitarian theory of testimony the two crucial claims of his *Communitarian Epistemology* - that knowledge is a social status and that the granting of this status depends on communities - are also further elucidated.

A Communitarian Account of Testimony

That Kusch starts the development of his *Communitarian Epistemology* with the topic of *testimony* is indicative of the central role this topic has had for the development of social epistemology in general. Indeed, for the longest time *testimony* has been used as a covering term for all social aspects of knowledge within epistemological discourse. Analyzing the sociality of knowledge *was* investigating the role of testimony for knowledge understood as the *transmission of knowledge from one individual to another*. But, according to Kusch, the sociality of knowledge goes far beyond a mere *transmission of knowledge* from one person to another. To vindicate a communitarian account of testimony, Kusch starts by exposing the limits of individualistic understandings of testimony. He argues that the scope of testimony was too narrowly conceived and that the notion of *epistemic interdependence* might be a better concept to describe the interrelatedness of epistemic agents. Moreover, the idea that testimony is a mere means of knowledge transmission has to be replaced with the acknowledgement that testimony often is a *generative source of knowledge*

Critique of Individualistic Accounts of Testimony

According to Kusch, there are two crucial flaws in individualist accounts of testimony (Kusch 2002). First, it has often been mistakenly assumed that testimony exhausts all social aspects of knowledge. Secondly, testimony was only considered to be a mechanism for the *spread of existing knowledge,* for the transmission of knowledge from one person to another. This distortion becomes understandable, when taking a look at the usage and the connotations of the word "testimony" in everyday life. "Testimony" is closely linked to eye witness reports in legal settings and this is the context in which it probably is most frequently used. Using this label "testimony" to account for the social aspects of knowledge has led to a framing and narrowing of the sociality of knowledge to cases similar to this legal setting, where one person is reporting what she has seen or what she knows. Thus, unreflective usage of the label "testimony" has had probably unintended, but detrimental effects on understanding the sociality of knowledge within mainstream (social) epistemology.

What Kusch finds particularly worrisome is that with such a narrow view on testimony the possibility that testimony could be more than just transmission of existing knowledge, but also a means of generating new knowledge has been neglected, or even rejected. To outline, in which ways testimony is generative of new knowledge, Kusch develops a concept of *performative testimony*. This concept is based on the assumption that performative speech acts are important sources of new knowledge and as such have to be accounted for by any comprehensive theory of testimony. His *Communitarian Epistemology* of testimony aims at being such a comprehensive theory of testimony and I depict it below in more detail.

Two topics are central in current debates around the epistemology of testimony: inferentialism and reductionism. Inferentialism concerns the question of whether our adoption of beliefs based on testimony is or should be direct or inferential; whether our adoption or rejection of testimony should be based on assessment of the testifier's competency and honesty. Kusch argues that the debates in this field have been impaired by a confusion of different levels of analysis. Phenomenological and pragmatic questions around testimony have been interspersed with psychological and normative questions. To his mind, much of the debates around inferentialism have not been particularly enlightening with respect to an understanding of the epistemological function of testimony. He concludes that "[a]n epistemology of testimony worth its salt had better avoid engaging in psychological speculations about subconscious processes; it had better stay clear of the misbegotten project of a phenomenology of testimony; and it had better give proper heed to the diversity of social practices surrounding the giving and receiving of testimony" (Kusch 2002: 28).

The second major topic in epistemological accounts of testimony besides inferentialism concerns the general justification of testimony. The basic question is whether testimonial knowledge has to be justified by other, *more fundamental* sources of knowledge, such as memory, perception or inference. Kusch argues that much of the debate around whether testimony has to be justified by other epistemic sources again is a result of the legal connotations of the term *testimony*. In court, distinguishing testimony based on former empirical observation from mere hear-say is clearly necessary. However, once one leaves the courtroom, the situation changes. If one adopts a view of *epistemic interdependence*, i.e. of the interrelatedness of all epistemic sources, the reduction of testimony to *more fundamental* sources becomes implausible. If memory, perception, inference and testimony are intrinsically interrelated, no source is more fundamental than the other.

After analyzing different forms of reductionism and anti-reductionism, Kusch argues that so far no convincing argument for a general justification of our reliance has been delivered (Kusch 2002: 43). Therefore, Kusch ends up arguing for communitarian quietism and contextualism and summarizes this position and its consequences as follows:"Since testimony is constitutive of all forms of justification, it is senseless to provide a general justification for it. Only local and contextual justifications are meaningful" (Kusch 2002: 77).[86]

[86] The interdependence of epistemic sources as well as the consequences of Kusch's contextualist position are addressed in more detail in the next chapter.

Predecessors of Kusch's Communitarian Account of Testimony

There are two communitarian accounts of testimony that Kusch considers particularly inspiring for his own approach: John Hardwig's work on the role of trust in knowledge (Hardwig 1985; Hardwig 1991) and Michael Welbourne's communitarian account of testimony developed in "Community of Knowledge" (Welbourne 1986). I depict these approaches here only briefly and in so far as they are relevant for Kusch's account of testimony.

John Hardwig's Role of Trust in Knowledge

In his paper "The role of trust in knowledge", Hardwig raises two issues which are highly relevant for any social epistemology and to which I return in the next chapter in more detail (Hardwig 1991). One topic concerns the distribution of epistemic labor, the other concerns the question of epistemic subjects in science. In contrast to Miriam Solomon (Solomon 2001), focuses on the distribution of work *within research groups* and not between them, i.e. within scientific communities as a whole. Arguing that the majority of research is nowadays conducted in teams, Hardwig emphasizes the fundamental role of *trust* in other epistemic agents for research. The second topic raised in his paper concerns the entity of the epistemic subject in science and is of primary concern for Martin Kusch's communitarian account of testimony. This issue centers around the question: Who knows? And more specifically: Who knows, if scientific knowledge has been created collaboratively?

Hardwig offers three possible answers to this question and uses a case study from quantum physics to outline the problem. In this example, 99 physicists co-authored a paper on charm particles and took part in the creation of the scientific knowledge depicted in this article. No single researcher could have reached the collective result on her own. Hence, the question is: who knows this knowledge? If we demand that knowledge is only possessed by individuals and that in order to know, one has to be able to provide evidence for this knowledge, then in the case of the physics example no-one knows, because no single scientist can provide evidence for all aspects that were necessary for this communal knowledge creation. Hardwig (Hardwig 1991) and Kusch (Kusch 2002) assert that this conclusions in not quite intuitive since it would imply that most scientific knowledge is not known by anyone. Thus, if we want to maintain the individual as the knower and consider such scientific knowledge to be known, we have to drop the requirement that knowers have to be able to marshal evidence in order to know. And such a relaxation of requirements might not be acceptable for many analytic philosophers.[87]

The alternative that Kusch proposes and that he also ascribes to Hardwig consists in taking the *community as a whole to be the primary knower*. Adopting such a communitarian view ”[...] allows us to retain the idea that a knower must be in ‘direct’ possession of the evidence but it breaks with the assumption that such a knower must be, or can be, an individual” (Kusch 2002: 49). Thus, Kusch interprets Hardwig’s analyses as an argument for communitarism, because it solves the dilemma of either having to refrain from ascribing knowledge to anyone or to drop the requirement of justification for the scientific examples provided by Hardwig (Hardwig 1991).

Moreover, Kusch also considers Hardwig’s analyses to provide an argument for the *generative role of testimony*. He argues that “given the fast and endless cycles of discussion and information exchange in, say, a place like CERN, the thesis that testimony is not generative loses all plausibility. Reports coming from other teams, the work of one’s own primary team, and finally one’s own work (usually itself part of a joint action with someone else) are so tightly interwoven that it is impossible to say where others’ input ends and one’s own processing starts” (Kusch 2002: 52).

Michael Welbourne’s Community of Knowledge

The second theory Kusch regards as a predecessor of his work is Michael Welbourne’s communitarian account of testimony as developed in “The Community of Knowledge” (Welbourne 1986). Welbourne’s arguments rest on the idea that knowledge can be communicated through say-so to a *believing hearer*. His theory is based on a rehabilitation of a common sense understanding of knowledge which he sets against mainstream epistemology. His critical stance towards mainstream epistemology becomes quite obvious in the following quote. “The philosophy of knowledge”, he argues, “has suffered greatly [...] from the obsession of philosophers with the word know” (Welbourne 1986: 3). And arguing for a more practice-related view on knowledge, he continues: “Understanding knowledge is primarily a matter of understanding the role which the concept plays in the life and conversation of human

[87] Longino (2002c) has proposed a solution to Hardwig’s dilemma that is depicted in the next chapter.

beings, the practices which it dominates. This understanding will not be gained just by puzzling out truth-conditions for sentences of the form 'a knows that p'; still less will it be gained by reflecting on the conditions, too often mistaken for truth-conditions, for responsible first-person singular assertions that one knows. One will get much further by examining the practice of telling the facts." (Welbourne 1986: 3).

There are three aspects that Kusch stresses in his reference to Welbourne: the understanding of knowledge as being commonable; the untypical role of belief for knowledge; and the central role of the community for knowledge. First of all, Welbourne argues that "[k]nowledge is essentially *commonable*. That is knowledge, by its very nature it can be made the common possession of two or more people by simple say-so, written, spoken or, in suitable contexts, gestural" (Welbourne 1986: 1). Further, when knowledge is transmitted, it is not removed from its original place nor does it lose quality (Kusch 2002). To use a more economic term, knowledge is non-rival (Foray 2004: 94). Not only can it be used individually as often as wanted without being worn-out or exhausted. It can also be used by a multitude of people without losing its value or quality. This aspect becomes crucial also when analyzing socio-epistemic practices on the Web in Part 3 of this thesis.

Secondly, the role of belief in Welbourne's approach differs profoundly from many traditional epistemologies, in which *knowledge is considered to be a certain kind of belief*, mostly *justified true belief*, sometimes with some additional requirement. Welbourne argues that knowledge is *not* a type of belief and that knowledge transmission is entirely different from belief transmission. However, belief becomes relevant as an *attitude of the recipient of testimony*. Knowledge transmission in this account is based on three prerequisites: that the speaker has the knowledge and is honest, that she communicates this knowledge via an appropriate speech act and finally that the speaker believes the testifier (Kusch 2002: 54). Thus, in this account, the attitude of the recipient of testimony is crucial, because the willingness to believe the testifier is a necessary prerequisite for a transfer of knowledge to occur. It is not the assessment of the speaker's competency or honesty that is crucial, "[a]*ll* that is required of a listener who understands a knowledgeable teller if the knowledge is to be successfully transmitted to him is that he *believe* the teller. (Welbourne 1986: 5f)

The third aspect concerns the role of the community for knowledge in Welbourne's account. That the community is quite crucial becomes obvious in the title of his book, but *how* exactly is the community related to the transmission of knowledge? According to Welbourne, when someone tells another person something and is believed, the two create a *primitive community of knowledge*. Such a primitive community "[...] consists of two people knowing the same thing and recognizing each other as sharers in that knowledge; so each can act on the assumption of knowledge in the other and they must be able to act co-operatively" (Welbourne 1986: 25). This community is dynamic in the sense that is depends on an act of communication that constituted it.

Kusch uses Welbourne's approach to make two claims that go beyond Welbourne's own conclusions and prepare his own approach. First, in combining Welbourne's insights with Brandom's (Brandom 1994) notion of *entitlement and commitments,* he concludes that "[...] sharing knowledge with others amounts to sharing entitlements and commitments with them" (Kusch 2002: 59). Referring to Welbourne's claim (Welbourne 1986: 84) that is it useless to insist on the value of a £10 note in a community that has no use for this note, Kusch (Kusch 2002: 62) further asserts that "[...] 'knowledge' is a social status like money, and thus it only exists in so far as there are items upon which we are willing to impose the status". And as noted before, it is always a community that has to grant this status.

Such a central role of the community for knowledge has implications for questions of objectivity in communitarian epistemologies. If knowledge depends on communities – does this dependence on the community diminish the objectivity of knowledge? Quite to the contrary, Welbourne argues. Instead of relativizing the possibility of objectivity, such a primitive community for knowledge actually *enables* objectivity in the first place. He states that the primitive community "[...] essentially [is] the product of communication: it is informed by the idea of commonable knowledge and a sense of reciprocity, of mutual support and dependence, which will seem to confirm each member in the knowledge which he has. I shall argue later [...] it is only in so far as Jones and Smith can think of themselves as sharing the knowledge that p that they can have the idea of an objective matter of fact at all."(Welbourne 1986: 26).[88]

[88] That the community is the basis for objectivity is also characteristic of Helen Longino's (2002c) *Critical Contextual Empiricism* which is portrayed below.

Kusch concludes that while both, Hardwig and Welbourne deliver crucial insights for a communitarian account of testimony, they do not go far enough. While John Hardwig (Hardwig 1991) focuses on the *object of knowledge*, on how evidence can be pooled so that a group of people can know more or things that they could not know individually. Welbourne's Welbourne 1986 emphasis lies on the *subjects of knowledge*, i.e. on the question of how testimony constitutes knowers with specific entitlements and commitments. Kusch argues that for a *Communitarian Epistemology* both aspects need to be combined. More specifically, he asserts that it is the *same process* that generates new communal *subjects* of knowledge and new *objects* of knowledge. His communitarian account of testimony is an attempt to make explicit how testimony is related to the generation of knowledge as a social status in general and how it is imposed on people in specific situations as is outlined in the next section.

Kusch's Communitarian Account of Testimony

Kusch develops his communitarian theory of testimony in five steps. The first step consists in introducing his central theoretical innovation: the concept of performative testimony. This concept of performative testimony is derived from the notion of performative speech acts. Performative speech acts are sentences such as "I hereby declare you husband and wife". Kusch argues that since performative speech acts are as much part of our communicative daily life as constative speech acts, they should be accounted for in any comprehensive theory of testimony. Moreover, it is through such performative speech acts that we impose social statuses on others. And given that Kusch considers *knowledge to be a social status*, the centrality of performative speech acts is not just a matter of complementing an analysis of constative testimony.

Performative speech acts differ from constative speech acts in their world-language relationship. While for constantive speech acts, the goal of language is to *match* the world, performative speech acts rather *change* the world. Thus, the direction of fit is different: for performative speech act, the world is to fit language, while for constantive speech act, utterances aim at fitting the world. By uttering a sentence such as "I hereby declare you husband and wife" – under the right conditions, *the world is made to fit this utterance*. In Kusch's terms, performatives are self-referring and self-validating, i.e. not only refers a performative utterance to itself, it also validates itself by transforming the world according to its content. Due to these two features, performative speech acts establish new social facts, create new knowledge for the listeners, and are as such a *generative source of new knowledge* (Kusch 2002 17). Kusch concludes that performative testimony is "[...] an important generative source of knowledge: the say-so constitutes a social fact, and it does so by creating a new item of knowledge for the couple, the witness, and the registrar herself (Kusch 2002: 17).

In the second step, Kusch introduces the notion of *communal performative testimony*, indicating utterances done by a 'collective we'. In the case of marriage it becomes obvious that not everyone is allowed to impose the social status of being married upon a couple, but that this person has to have been granted the status of being in the position to grant the status of being married himself. These rights to grant status are typically granted by social institutions. Kusch uses a very broad notion of 'social institution' here and includes conventions such as greeting others under this term. Kusch argues that these *social institutions themselves originate in performative speech acts conducted by a community*, such as "[w]e hereby declare that it is correct to greet people one knows" (Kusch 2002: 68). Obviously, this statement has never been uttered by a Greek chorus to come into effect. Kusch argues that instead "[…] communal institution-creating performative testimony *is typically fragmented and widely distributed over other speech-acts* [italics in the original]. The communal performative is never explicitly made; it is only made implicitly or indirectly. It is carried out by people when they *do other things* [set in italics by JS]" (Kusch 2002: 67).

I have set this last phrase into italics, because it introduces the third step of Kusch's argument. He states that almost all testimony is in part performative and as such generative of new knowledge. Thus, the social institution of greeting others, gets constituted by the *practice of greetings others* and not by a collective announcement of the necessity to greet others.

Thus, in the third step Kusch shows that performative and constative testimony is indeed deeply intertwined. The relevance of performative testimony is most obvious when considering *social kinds*, such as money or marriage. The antonyms of social kinds are natural and artifical kinds. And this differentiation between social, natural and artificial kinds refers to the effect that the withdrawal of a communal performative testimony has on them. To elucidate this difference Kusch uses the three examples of an elephant (natural kind), marriage (social kind) and typewriter (artificial kind). While an elephant and the typewriter continue to exist, although not *as* elephants and typewriter when we withdraw the communal performative that defines what an elephant or a typewriter is, *social kinds, such as money or marriage cease to exist the moment, the communal performative testimony is withdrawn.* Take again the example of marriage: without the performative act of declaring someone husband and wife, there is no independent reality of a married couple. The moment the communal performative testimony is withdrawn - let's say it gets known that the alleged priest is a fraudster – the marriage ceases to exist.

Although the impact on social kinds is clearly more fundamental, performative testimony nonetheless plays a role also for natural kinds (e.g. elefants) and artificial kinds (e.g. typewriters). "Every constative testimony about elephants carries part of the communal performative speech-act which constitutes the category of elephants, and in doing so, re-enforces the conventional ways of delimiting this category, and helps to entrench the conventional exemplars" (Kusch 2002: 69). Thus, a mutual relationship between performatives and constatives can be observed: "On the one hand, constatives presuppose communal performatives. We can make claims about marriages, elephants, or typewriters only because our communal performatives have constituted the taxonomies and exemplars needed for making such claims. On the other hand, constatives perform (i.e. partially constitute) communal performatives. They carry out the fragments of widely distributed communal performatives" (Kusch 2002: 69).

In step 4 Kusch identifies *knowledge as a social kind* as compared to natural or artificial kinds and concludes that generation of knowledge takes place in a quite similar form as the creation of greeting as a social institution. Accordingly, attribution of knowledge not only depends on these distributed and fragmented speech-acts, but it is also embedded in other practices. And if knowledge is a social kind, as Kusch argues, it also ceases to exist the moment it is not communally considered to be knowledge anymore. Within such a general approach, the question remains open how we decided about knowledge claims in specific situations. Just stating that knowledge comes into being by these distributed, communal acts does not help much in deciding what we take to be knowledge and what we reject. How do we decide whether something is knowledge or not? How do we decide whether to rely on someone's claims? How do we decide whom and what to trust when searching for knowledge?

Unsurprisingly, the answer to this again lies in the community. The last step in Kusch's development of his communitarian theory of testimony hence consists in emphasizing the relevance of epistemic communities for the attribution of knowledge. Kusch states that something that is testified can acquire the status of knowledge only if an epistemic community is formed around it. For something to become knowledge, a recipient of testimony has to enter into such as relationship with the testifier. Before this community formation knowledge does not exist. Hence, the testifier does not *possess* knowledge that he just *transfers*. An individual cannot possess knowledge in total solitude, because as a social status this privilege can only be granted by a community. The individual might *believe* something, or be convinced of something – as long as a community of a minimum of two people does not agree on this, there is no knowledge. Quite interestingly, Kusch considers teaching, which could be considered to be the prime example of testimony, to be an *exception* of this form of testimony (Kusch 2002: 73). In the case of teaching, the knowledge was generated *before* being transmitted to the student, who has not been involved in the initial act of constituting this knowledge. That is, in the case of teaching 'p as a fact', where pure knowledge transfer takes place, the relevant epistemic community was formed *before* this act of transmission and with other people who formed an epistemic community to create this knowledge.

5.4.5 Helen Longino's Critical Contextual Empiricism

Introduction

The final comprehensive social epistemology that I portray is also the most instrumental one for this thesis: Helen Longino's *Critical Contextual Empiricism*. In her book "The fate of knowledge", Longino develops an account of scientific knowledge that is supposed to be "[…] responsive to the normative uses of the term "knowledge" *and* to the social conditions in which scientific knowledge is produced." (Longino 2002c: 1).[89] She argues that there have been two major shifts in the current philosophy of science, namely a growing recognition of the social character of science on the one hand and of explanatory plurality in science on the other. With her approach, which she labels *Critical Contextual Empiricism*, she aims to take into account insights on the nature of scientific inquiry obtained in the social studies of science, i.e. the various *empirical* approaches to understand scientific knowledge creation while sticking to the *normative*

[89] My depiction of her approach is primarily based on this book unless otherwise noted.

aspirations of epistemology. Longino focuses primarily on knowledge creation in the sciences, taking into account literature from sociology and philosophy of science. Nonetheless, I think that her approach does not need to be restricted to knowledge creation in science only. Rather, her observations with respect to plurality and provisionality as well as her recommendations regarding proper forms of criticism can easily be applied to knowledge creation in more informal contexts.

Dissolving the Rational-Social Dichotomy

Longino's initial target of criticism is what she calls the "social-rational dichotomy" (Longino 2002c: 1). She argues that both sociologists and philosophers of science have fallen victim to this "dichotomy between the rational (or cognitive) on the one hand and the social on the other [which] structures both (1) the disagreements between the practitioners of the social and cultural studies of science and the philosophers and (2) the constructive (or deconstructive) accounts they all offer of scientific knowledge" (Longino 2002c: 1).

To develop her own account, she first analyzes some of the contributions of the social studies of science for an understanding of scientific reasoning. More specifically, she critically assesses the works of the Edinburgh school and the laboratory studies conducted by Bruno Latour and Karin Knorr-Cetina. While appraising the insights obtained on the actual practices in the sciences, she argues that the irrelevance of normative philosophical concerns proclaimed in these approaches can be explained by their adherence to the social-rational dichotomy and should be overcome. She then continues to analyze the philosophers' responses to the these sociological challenges, arguing that the – often strong - reactions of philosophers such as Alvin Goldman, Philipp Kitcher or Larry Laudan can also be explained by their entanglement in the rational-social dichotomy. Sociologists who deny that science has anything to do with rationality but rather is determined by social forces and or material conditions only commit the same mistake as philosophers who assume that since science is the most rational endeavor sociality can only come to the fore by being *detrimental* to rationality. Both versions are the flipsides of the same coin and bound to an antithetical conception of sociality and rationality: what is rational is by definition not social and what is social cannot be rational. Longino asserts that as long as this dichotomy is not resolved, neither can there be a rapprochement between sociologists and philosophers of science, nor can science be properly understood as a rational *and* social practice. Accordingly, Longino's goal is to disentangle the rational-social dichotomy to open up a new line of reasoning about scientific knowledge creation that can account for both: the *actual social practices of science* as well as the normative concerns of philosophers, i.e. the recognition of *knowledge as a success term* (Longino 2002c: 10).

Three Modalities of Knowledge

Longino pursues this goal by at first disambiguating three senses of knowledge: *knowledge as content*, knowledge as a set of *knowledge-productive practices* as well as *knowledge as cognitive agency*, i.e. as a state of a person ('knowing'). Since I adopt Longino's tripartite notion of knowledge as the framework for my own analyses, I devote some space to it here and return to it in the next chapters as well. Longino argues that not only do philosophers and sociologists often understand and emphasize these three notions quite differently. Moreover these three different senses of knowledge do also get frequently mixed up within different accounts on scientific knowledge creation. Longino concludes that the "failure to distinguish the different senses undermines efforts to make the case for the sociality of knowledge" (Longino 2002c: 77) and thus starts disentangling them.

Knowledge Production

Knowledge production deals with the transformation of various inputs, such as sense data into representational outputs, i.e. the question of how knowledge is produced. But while sociologists focus on the processes by which cognitive authority *is actually granted* to someone and how the legitimacy of knowledge claims gets decided, the philosophers note that neither cognitive authority granted nor the attribution of knowledge to someone might be *warranted* in every case. Hence they are rather interested in criteria to decide when the attribution of cognitive authority and legitimacy is *justified* and when it is not.

Cognitive Agency

Knowledge as knowing by contrast describes the state of a person towards some object and has been the main focus of attention in analytic epistemology, but has been of much lesser interest to sociologists. Knowledge in this sense is characterized as a three-term relation between a subject, an object and a representation, frequently associated with the notion of *knowledge as justified true belief* and the omnipresent proposition *"S knows that p"*.

Knowledge as Content

The last sense of knowledge Longino distinguishes is *knowledge as content*, i.e. knowledge as that which is known, the corpus of knowledge, its materialization and aggregation. Here again, philosophers and sociologists focus on different aspects of this modality of knowledge. While the sociologists are rather interested in what a given community *considers knowledge*, the philosopher searches for a *validation* of knowledge independent of any given community. The result of this search may indeed be a negative one: some authors, such as Martin Kusch and Helen Longino herself conclude that knowledge always depends on communities and thus cannot be conceived independent of them. That means while sociologists focus on the empirical aspects of how knowledge gets validated, while neglecting or rejecting any normative questions of whether these practices of validation are appropriate; philosophers have been preoccupied with normative issues of how such validation should take place while often being negligent of the empirical question of how knowledge actually gets justified.

Longino argues conclusively that many misunderstandings as well as much of the conflict between sociology and philosophy of science have their roots in a lack of recognition of the different senses of knowledge referred to as well as in a constant blurring of normative and descriptive foci. While acknowledging that these three senses are interrelated, Longino nonetheless argues for an analytical separation of these concepts to clarify the debates. She summarizes these differences in the following table (Longino 2002c: 84).[90]

[90] Explanation of abbreviations: PP_e = knowledge productive practices, empirical, PP_{n} = knowledge productive practices, normative, C = community, S = subject, *p* = proposition.

	Empirical	Normative
Knowledge-productive practices	PP_e: processes or practices that succeed in fixing belief or in having some content accepted in some community	PP_n: processes or practices of belief acquisition that justify belief
Knowing	S accepts that *p*, and *p* is accepted in C, and S's acceptance of *p* is acceptable in C	S accepts that *p*, and *p* is true, and S's accepting that *p* is the outcome of or accords with PP_n
Content	$Content_e$: what is accepted in some community C or the outcomes of PP_e in community C	$Content_n$: the subset of truths which is known (whether by an individual or by a community)

Table 2: Longino's Three Senses of Knowledge

The Sociality of the Three Senses of Knowledge

Having distinguished these three senses of knowledge, their distinct comprehension as well as the different foci in philosophy and sociology, Helen Longino continues to show in which ways and to what extent all three senses are social. Socializing the three modalities of knowledge lead her to the following conclusions: knowledge-productive practices and their modes of justification vary in different contexts, cognitive agents are interdependent and knowledge as content is plural (Longino 2002c: 122).

The Sociality of Knowledge Producing Practices

Longino starts with arguing for the sociality of knowledge-productive practices, i.e. cognitive or intellectual as well as the material practices involved in science. The two major examples she chooses are *observation* and *reasoning*, because both are basic epistemic practices in science.[91] Opposing epistemic individualism, Logino proposes "[…] to treat both observation and reasoning as dialogical, that is, as activities involving discursive interactions among different voices" (Longino 2002c: 99). Due to an overly individualistic philosophical tradition both practices have been considered to be private endeavors for the longest time. However, observation regarded as "pure perception" has been the first primary source of knowledge whose alleged solitude has been disenchanted.

[91] Longino also considers material interventions and statistical procedures to be epistemic practices. Since socio-technical epistemic practices as well as stochastical and other mathematical procedures are an important aspect of many Web2.0 applications such a broad view on epistemic practices is highly relevant for my analyses.

Observation

Analyses of science conducted in anthropological, sociological and historical analyses of science , such as Shapin's *A Social History of Truth"* Shapin 1994, have shown that scientific observation is by no means *pure* sensory perception, whatever that may be, but rather a process of ordering and organizing observational data, a public calibration of perception, and as such a *social* matter.[92] The organization of these data is dependent on categories, classes and the boundaries between them as well as on theoretical and methodological background assumptions. Such background assumptions structure not only the organization of data, but also their very way of elicitation from the beginning. Thus, while sensory perception might be an individual act, observation can only be achieved collaboratively.[93] It is not only reliant on consensual acceptance of central categories and background assumptions. Social mechanisms are also "[…] what enables the transformation in assertability-status from "It seems to me that p" to "P"." (Longino 2002c: 103). Only through interactive processes can perception became observation and hence opinion turn into knowledge.[94]

It is in this fundamental sense, that knowledge in the sense of knowledge-productive practices is *social* in Logino's approach. She makes very explicit though, that the term *social* in her account is different from *common, collective,* or *shared*, but rather should be understood as *interactive*. Instead of sharing epistemic goals or having common knowledge, the crucial sociality of knowledge lies in being in a *dialogue* about issues at stake; it is about interacting in producing situated, partial and provisional knowledge.

Reasoning

That epistemic practices are social holds true not only for observation, but also for scientific reasoning, an argument which at first sight may appear less plausible. Most people would think that reasoning - maybe sitting in your armchair with a pot of tea, pondering on some question - is the most individual practice one can imagine. And indeed, Longino states that while "[b]oth inference and observation are mediated, […]

[92] Shapin's work will be portrayed in more detail in the next chapter.
[93] I would argue that even this notion has to be rejected since perception is influenced by learned categories and by higher cognitive processes.
[94] This dependence of knowledge on the interaction between at least two persons is also characteristic for Kusch's (2002) and Welbourne's (1986) communitarian epistemologies.

the degrees of freedom in inference or reasoning are greater than in observation" (Longino 2002c: 104, footnote).

Reasoning, Longino argues, is more than just calculation and has two meanings, a *constructive* and a *justificatory* one. While in the constructive sense "[...] reasoning is the combining of ideas or information to *produce* new ideas", in the justificatory sense, "reasoning is the combining of ideas or information to *support* some other idea" (Longino 2002c: 103, emphasis added). Thus, Longino differentiates between creative reasoning and supportive reasoning, i.e. between the generation of new ideas and the confirmation of such ideas.[95] While the cognitive abilities of individuals are "[...] both necessary and sufficient for the generation of ideas" (Longino 2002c: 122), it is in the justificatory part that the social processes of reasoning becomes obvious. Justification according to Longino is a social practice and its standards are determined through social interaction. Since knowledge depends on such a social process of validation "[…] what counts as an appropriate consideration, as a reason, is determined and stabilized through discursive interactions" (Longino 2002c: 103f). Justification of knowledge is social in a double sense: not only is the *mechanism* of justification social by involving others, its *standards* are also discursively enacted.

Thus, concerning the sociality of epistemic practices, Longino concludes that "[...] discursive interactions are integral to both observation and reasoning in the sciences. The results of both reasoning and observation, then, are socially processed before incorporation into the body of ideas ratified for circulation and use, or are treated as having been so processed" (Longino 2002c: 106). Acknowledging the social nature of justification is consequential. If these standards are socially negotiated, how can we make sure that they are not completely wrong or biased? Longino's solution consists in requiring the participation of "[...] multiple points of view to insure that the hypotheses accepted by a community do not represent someone's idiosyncratic interpretation of observational or experimental data" (Longino 2002c: 106). Together with the requirement of *empirical adequacy*, this is all that can be done to avoide falsity and bias.

[95] For philosophers of science this distinction between a creative and a justificatory component of reasoning might sound familiar and reminiscent of Reichenbach's differentiation between the context of discovery and context of justification (Reichenbach 1938).

Two aspects are crucial concerning her request. First of all, directly after having accounted for the sociality of scientific practices, Longino draws *normative* conclusions about the impact this observation *should* have on epistemic practices and the structure of science. And to my mind, this constant *interrelation of descriptive with normative aspects* is one of the major merits of Longino's approach. Secondly, Longino is very aware that the sociality of knowledge is a two-edged sword. Being a feminist philosopher, Longino is far from stating that the sociality of knowledge is without problems or cannot be detrimental to science. Biases and discrimination as social injustices can play a role in science as much as they do in other societal domains. However, she argues that their sociality is not only a source of problems, but can also be a solution to them. Seen in this respect; "[…] socializing cognition is not a corruption or displacement of the rational but a vehicle of its performance" (Longino 2002c: 106). Thus, the social nature of epistemic practices such as observation and reasoning comes with pros and cons and Longino's request to include multiple voices into scientific practices is exactly an example of using social mechanisms to counter social injustices, biases and corruptions. The sociality of knowledge-productive practices can be a source of error and bias as much as a cure for them. But either way, it is unavoidable and hence has to be accounted for – descriptively and normatively.

The Sociality of Cognitive Agency

Longino continues to show then that not only are these knowledge-productive practices social, but that the same holds true for *knowledge as agency*. Acknowledging the sociality of cognitive agency for Helen Longino means to reject the unconditioned subject S, known from the classical "S knows that p". Again alluding to insights from sociology, she argues that "[…] subjects creating scientific knowledge are located— historically, geographically, socially— and that their locatedness must be taken seriously. To acknowledge the locatedness of subjects is to reject what we might call an "unconditioned subject," that is, a knower guided only by context-independent and value-neutral methodological rules (Longino 2002c: 107)." And instead of being independent - or fully determined, *knowers are interdependent of each other* and it is in this sense, that knowledge as cognitive agency is social.[96] Longino therefore stresses two different but related social aspects of knowledge as cognitive agency: Not only is

[96] This acknowledgement of the interdependence of epistemic agents resonates with Kusch's (2002) emphasis of the epistemic interdependence as a replacement for the concept of testimony.

the knower a subject situated in a certain social environment and shaped by her experiences in this environment. She is also connected to other knowers through her interactions within her cognitive communities.

The Sociality of Knowledge as Content & the Notion of Conformation

Finally, Longino concludes, that knowledge in the third sense, i.e. *knowledge as content* is social, too. She argues for the *plurality of knowledge* which can be addressed through the notion of *conformation.* Two aspects of this third modality of knowledge, knowledge as content, are important. First of all, *knowledge* in this sense is considered to be more than just mental, propositional content. It is something that can be stored and transmitted – in documents as well as in memory. Moreover, knowledge can not only be embodied in text, but also pictures, diagrams and other non-textual formats. Further, knowledge also includes what Longino labels *practical knowledge,* and is otherwise referred to as *tacit knowledge* (Polanyi 1985) or *procedural knowledge.* Such a broad notion of knowledge is necessary for the analyses of epistemic social software as is shown in the third part of this thesis.

Secondly, knowledge is used as a *success term,* i.e. it is meant to distinguish successful content (= knowledge) from unsuccessful content. It is essentially this second characteristic that often divides the philosophers from the sociologists. While success of content for the sociologist denotes that which *gets accepted* in a certain community, success for the philosopher usually has to do with a *truth* or *correspondence to reality*. Based on analyses by Nancy Cartwright (Cartwright 1983) and Ian Hacking (Hacking 1992a) concerning the notion of truth in the science and in line with model-theoretical accounts of scientific theories, Longino denies that truth is a good indicator for the successfulness of content in science. As was also argued by Miriam Solomon (Solomon 2001), most theories in the history of science are strictly speaking not *true*. Instead of adopting Solomon's *Whig realism* in order to keep the notion of truth, Longino argues differently. She proposes her notion of *conformation* as an alternative to the true-false-distinction (Longino 2002c: 115ff).

Conformation is meant to be an umbrella term for the empirical success of content. Notions of truth, isomorphism, homomorphism, fit, similarity or approximation are just special instantiations of it. Conformation is dependent on the *purpose* of an epistemic task. And as compared to the true-false-dichotomy, it comes in degrees. Comparing theories with maps, Longino exemplifies her notion of conformity very clearly. Theories, just as maps, are made for certain purposes and might serve these purposes better or worse. Of course, in order to be useful, maps have to be empirically adequate, but maps are hardly *true*. What should the *truth* of a map be? Being a duplicate of the terrain they are intended to map? Quite to the contrary, it is a crucial characteristic of maps, that they *do not* just duplicate the terrain they are supposed to map. Doing this, they would be useless. In a similar vein, theories are not meant to just copy the reality they seek to explain and as such are never strictly speaking *true,* at least not in the sense of corresponding to reality.

Moreover, comparing theories to maps also supports Longino's proposed *plurality of knowledge*. Since there can be different maps of the same terrain that serve different purposes, there is not one *true* map of this terrain or even the *best* map. Rather, depending on the question, one might make use of a political map or a geographical map, depending on whether I drive to place in France by car or want to go hiking in the Mercantour National Park I chose maps of different scales. Depending on my interest, a geographic or a political map of Germany may be better. Clearly, none of the maps of the maps is *false.* None of them even is essentially *better* than the other. Only for and within each purpose can I compare the quality of these maps.

However, such a context- and purpose-dependence understanding of empirical success does not imply that success is free choice and construction is just making things up. *Empirical adequacy* always also has to be given and Logino concludes:

> "Success can not just be a matter of the user wanting the theory to be correct or wanting to be able to act as if the supposed consequences of the theory are true. If they are not, reality will eventually bite back. To suppose that success has nothing to do with the interests of the users is just as fruitless, however. The idiom of representation must be such as to enable successful interaction with that which is represented. The choice of idiom, and of the degrees to and respects in which it must fit the objects of representation, is a social choice, a matter of goals collectively endorsed in the community conducting inquiry. In this sense the community determines what will count as knowledge." (Longino 2002c: 119).

A final advantage of conformity over truth is that it can be applied to non-propositional content as well. If one grants that knowledge comes in many non-propositional forms as well, i.e. in pictures, in diagrams, etc., then such a broader concept to assess the quality of different knowledge types should clearly be favored.

Critical Contextual Empiricism: Longino's Social Account of Knowledge

Helen Longino develops her social theory of knowledge based on the distinctions fleshed out before. She argues that her account is superior to other accounts on the sociality of knowledge because it not only meets the challenges posed by philosophers and sociologist. It can also handle the problem of underdetermination, i.e. the gap between observational data and theories, which has haunted philosophy of science for a long time. Her definition of *knowledge as epistemically acceptable content* then reads as follows:

"Some content A is *epistemically acceptable* in community C at time t if A is or is supported by data d evident to C at *t* in light of reasoning and background assumptions which have survived critical scrutiny from as many perspectives as are available to C at *t*, and C is characterized by venues for criticism, uptake of criticism, public standards, and tempered equality of intellectual authority" (Longino 2002c: 135).

The concept of *epistemic acceptability* forms the second key normative concept in Longino's approach together with the previously introduced notion of *conformation*. Before I explain it in more detail, I outline Longino's four social norms featured in the definition above: venues for and uptake of criticism, public standards and tempered equality of intellectual authority.

Social Norms for Knowledge: Four Norms for Effective Transformative Criticism

These four social norms mark the normative-prescriptive core of Longino's social epistemology and are also closely related to her feminist concerns (Longino 1996, Keller and Longino 1996). If one acknowledges that knowledge creation in science and elsewhere is a social process one also has to deal with the possibility of social inequalities and epistemic injustices within such processes. Therefore, a normative social epistemology has to provide guidance as to how to avoid or counter such biases. Longino's request to include as many perspectives as possible into the process of knowledge creation is an attempt to counter such bias. However just including different perspectives is not sufficient.

For instance, if certain people can participate but are not taken serious, then such an "inclusion" of multiple voices would be pure lip service. On the other hand, if everybody has a say irrespective of whether he actually knows anything about the topic or even tries to torpedo the epistemic endeavor, this can also not be beneficial for science or knowledge creation more generally. Therefore, the inclusion of multiple perspectives needs to be specified. And Longino does this by proposing four social norms that should be fulfilled. If fulfilled they provide the basis for *transformative criticism* and ensure that knowledge can be generated effectively and with the least possible bias. What exactly are these four requirements that such an epistemically ideal community would have to satisfy?

The first requirement consists in the availability of "[…] publicly recognized forums for the criticism of evidence, of methods, and of assumptions and reasoning" (Longino 2002c: 129). Venues for criticism have to exist and they have to be publicly recognized - in other words: criticism needs more space and more recognition. Criticism not only advances research and thinking by identifying mistakes or biases. It also helps to understand the grounds, limits and consequences of knowledge to a greater extent, as Longino argues with reference to John Stuart Mill (Longino 2002c: 3f).

However, that people criticize others, even if this happens in the most constructive way, does not ensure that knowledge production is improved. The criticism might not be heard, ignored or even bluntly rejected without having been considered. Longino therefore demands as a second norms that it is not sufficient to just tolerate criticism, but requires that a community's "[…] beliefs and theories must change over time in response to the critical discourse taking place within it" (Longino 2002c: 129f). She concludes that only "[u]ptake is what makes criticism part of a constructive and justificatory practice" (Longino 2002c: 130). In contrast to the sheer availability of venues for criticism, this prerequisite requires the development of a new attitude; it involves learning and possibly pedagogic intervention to foster a more positive stance towards criticism.

The third critical factor in establishing an ideal epistemic community concerns the availability of public standards. Longino argues that for criticism to take place participants in such a dialogue must share some terms, standards and epistemic methods. In contrast to Longino's former usage of the term *social* as interactive instead of shared or common, in this case some degree of consensus seems to be necessary. In order to be in interaction, certain prerequisites, such as terminology and definitions, but also values of scientific conduct have to be shared and consented upon. Since these standards are related to the current goals of a community, they are not fixed for all times, but change with the community's development and its changing goals. Nonetheless, at least temporarily they have to be agreed upon. Moreover and similarly to Kusch's notion of 'communal performative testimony' (Kusch 2002: 66f), these standards are not generated by a public announcement or a single willful act, but rather they are developed and enacted while participants act according to them. Scientists co-create and re-enact these standards and values while performing and evaluating epistemic practices and results.

Longino's final requirement concerns *tempered equality of intellectual authority*. The request for equality is most directly linked to Longino's feminist orientation. However, this requirement is not meant to be merely a political, social or ethical imperative, but also essential for the effectiveness of knowledge production. She argues that "[t]he exclusion of women and members of certain racial minorities from scientific education and the scientific professions constitutes not only a social injustice but a cognitive failing" (Longino 2002c: 132). Longino backs this assumption with reference to various feminist analyses in the history and sociology of science, which have shown that scientific development has often suffered, because certain voices were not heard, because criticism by certain groups of people (women, racial minorities or scientists from different scientifically less acknowledged countries or institutions) were not taken serious and ignored (Longino 2002c: 132). Therefore, the feminist concern about the equality of different voices is not purely politically motivated, but also an epistemic duty. However, the word *tempered* provides a relativization of the proposed equality. Longino acknowledges that not everyone should be granted the same amount of authority on each topic. Further, it has to be avoided that the same criticism is uttered unchanged over and over again, i.e. the request to uptake criticism is also valid for the critic herself. Concerning the conditions of *how* to temper equality of intellectual authority, Longino does not give a definite answer, but confines herself to listing critical issues around it (Longino 2002c: 133f). I do agree that a general answer to the question of tempering equality cannot be given and that such decisions have to be made locally.

However, this issue surely is crucial to distinguish a reasonable selection of valid criticisms from epistemic injustice.[97]

Epistemic Acceptability: A Socio-Empirical Criterion for Knowledge

These four social norms then form the basis for Longino's definition of *knowledge as epistemically acceptable content*. However, they form *only one of its two parts as* can be seen from her definition of knowledge above.[98] Her emphasis on the necessity to provide empirical evidence to support knowledge claims shows that social mechanisms alone do not make knowledge claims epistemically acceptable – at least not in a normative sense.

Rather both, *empirical data and social mechanisms of validation and criticism* are needed for the transformation of mere belief into knowledge. Longino argues that while philosophers usually stress the underdetermination of theories, sociologist often commit a similar felony by arguing that theories are entirely determined by social mechanisms. Instead, she argues, both evidence (e.g. empirical or logical evidence) *and* social interactions play a role in the creation and justification of knowledge. Thus, her concept of epistemic acceptability offers a model of dual justification, of a social and empirical validation. In doing this, it unifies not only the traditional philosophical concept of justification by evidence with a social justification mechanism that works via public criticism, but also philosophical concerns about normativity with the sociologists' insights into the nature of knowledge production. She concludes that by this move, her account can not only avoid collapsing into relativism, while still being able to address the problem of underdetermination. The dichotomy between the rational and the social which has caused such a great deal of misunderstandings is also dissolved by combining the social and the rational as two equally important and related aspects of knowledge creation and justification.

To summarize, *epistemic acceptability* and *conformity* are the two central normative concepts in Helen Longino's social epistemology. While her notion of epistemic acceptability combines the classical philosophical concern with justification through

[97] The issue of how to temper intellectual authority will remain crucial throughout this thesis.

[98] "Some content A is *epistemically acceptable* in community C at time t if A is or is supported by data d evident to C at *t* in light of reasoning and background assumptions which have survived critical scrutiny from as many perspectives as are available to C at *t*, and C is characterized by venues for criticism, uptake of criticism, public standards, and tempered equality of intellectual authority." (Longino 2002c: 135)

empirical evidence with a social mechanism for effective criticism, conformity servers as an umbrella term for the relation to the aspect of reality a theory intends to map. Helen Longino's social and contextualist conceptions have consequences for the concept of knowledge they entail: Knowledge in this sense is initially always *partial, provisional and plural*, although these features might not be necessary on the long run.

6 Assessing Social Epistemology

6.1 Introduction

In the previous chapter I introduced the field of social epistemology and how I am conceiving it in this thesis, and portrayed five major socio-epistemological theories. I implied that the field of social epistemology has to be understood in a broad context of multiple, interdisciplinary approaches to understand the sociality of knowledge. Moreover, I argued that even in its narrow delineation social epistemology is not a unified field. Rather, different lines of conflict, different cleavages can be made out and the positioning of different authors along the lines has lead to various classifications of the field of social epistemology. For the most part, the field has been divided into two opposing positions; sometimes even tripartite notions of social epistemology have been proposed.

These classifications include the following confrontations:

- *analytic social epistemology* versus *critical social epistemology* (Quinton 2004)
- *descriptive* versus *revisionist social epistemologies* (Fuller 2004)
- *classical* versus *anti-classical social epistemologies* (Goldman 2006)
- *revisionist, preservationist* and *expansionist social* epistemologies (Goldman 2009b, Goldman 2009a).
- social epistemology as *individual doxastic agents with social evidence*, *collective doxastic agents* and *systems-oriented social epistemology* (Goldman 2009c).
- *complementary* versus *science policy programmes* versus communitarian epistemology (Kusch 2002),
- *diagnostic social epistemology* as a future discipline that could be differentiated into: analytic social epistemology, genealogical social epistemology, historical epistemology, political social epistemology and naturalistic epistemology. (Kusch 2009).

These attempts to classify the field of social epistemology are an indicator of the central debates within the field. From my reception of the various socio-epistemological theories, as well as from these proposed classifications, I have extracted the following points of conflict within the field:

- Complementarity: Is social epistemology complementary to some individual epistemology or does it replace individualistic epistemologies?
- What is knowledge and who is to decide? (Definition of knowledge, adherence or rejection of the definition of knowledge as justified true belief; stance towards sociological and historical approaches)
- The utility of truth: Is truth a useful and/or sufficient concept to assess epistemic practices and concepts?
- The relevance of the community: How central is the community? Implications concerning contextualism, local epistemologies and relativism.
- Methodological issues: Naturalism and the relevance of quantification.
- Normativity: normativity refers to two issues: a) normativity as providing criteria for knowledge and b) normativity in a prescriptive sense of making recommendations to improve epistemic practices and systems

I am by no means arguing that these are the only topics that are up for debate. However, they are surely central and important to consider when trying to understand conflicts within social epistemology debates. Thus, I use the arguments and the positioning of different social epistemologists along these lines of conflict as a grid for my critical analysis of the social epistemologies portrayed in the previous chapter. I do not aim at resolving the debates that I open up. The goal is rather to span the field of social epistemology by using the debates as vectors for analysis and to position myself in the field.

Although I have distinguished different debates, these are clearly not independent or unrelated. For instance, the utility of truth plays a role for the definition of knowledge, for normativity, for relativism, etc. Nonetheless, I think my breakdown of topics enables a clearer view on the central debates.

Before outlining the debates I want to focus on three core social mechanisms that were addressed in almost all of the major social epistemologies portrayed in the previous chapter: testimony, the distribution of cognitive labor and consensus formation. I return to these mechanisms in some detail and portray some additional perspectives on them for two reasons. First, I would argue that analyses on these three mechanisms form the core of the field of social epistemology as conceived in this thesis. Secondly, the socio-epistemological model that I develop in the next chapters provides a conceptual framework that is built upon these socio-epistemological mechanisms.

I conclude this chapter by linking social epistemology to social software. More specifically, I assess the utility that these five social epistemologies portray for analyses of socio-epistemic practices on the Web and outline some initial attempts to analyze ICT and social software from a socio-epistemological perspective. The analyses of this chapter then serve as a basis for the development of my theoretical framework for analyzing socio-epistemic practices on the Web – and beyond – in the succeeding chapters.

To summarize, I start this chapter by giving some details on the three central socio-epistemic mechanisms: testimony, distribution of cognitive labor and consensus formation. I follow by outlining the central topics in social epistemology along the lines of debate, and I conclude the chapter by linking to and assessing the merits and current limits of social epistemology for analyzing social software when it is understood from the vantage point of socio-technical epistemic systems. All of these analyses serve as the basis for the development if my own socio-epistemological model, which I develop in the next chapters.

6.2 Central Socio-Rational Mechanisms

Testimony, the distribution of cognitive labor and consensus are three central socio-epistemic topics and almost all social epistemologists have addressed them in their theories. In this section, I expand on these topics by providing information on the central debates and central approaches by theoreticians other than those introduced in the previous chapter.

In developing my own socio-epistemological model, I argue that testimony, the distribution of cognitive labor and consensus are three *socio-epistemic mechanisms* which denote central *temporal occurrences* in socio-epistemic processes. While distribution of cognitive labor refers to multiple ways in which socio-epistemic processes can be *started*, consensus is one distinct mechanism of *closing* such socio-epistemic processes amongst others. Testimony also refers to a central socio-epistemic mechanism, but for the most part this notion is exclusively focused on the *transmission of existing knowledge*, on the distribution of information to multiple receivers.[99] The

[99] Martin Kusch's (2002) notion of *performative testimony* has been a notable exception from this narrow focus on transmission. Nonetheless, the majority of analyses on testimony in epistemology still adhere to such a transmission model of testimony. While I agree with Kusch that such a narrow view on testimony

topic of *epistemic trust* is often discussed in the same context as testimony, because it refers to the stance that recipients of testimony have towards testifiers. However, as I argue in more detail in the third part of this thesis, trust refers to another dimension of epistemic sociality. While testimony, distribution of cognitive labor and consensus all refer to the *mechanisms of distribution and reuniting* either knowledge or epistemic tasks over different epistemic agents, trust refers to the *evaluation* of such agents. These two dimensions, distributing-reuniting work over epistemic agents on the one hand and evaluating them on the other, are crucial for the socio-epistemological model I outline in Part 3. First, however, I give some background on the central debates about testimony and trust, distribution of cognitive labor and consensus in socio-epistemological discourse.

6.2.1 Testimony

One of the most central topics in any social epistemology is testimony. Testimony, considered the fourth classical route to knowledge, in addition to *perception, inference* and *memory*, refers to the process of acquiring knowledge though the words of others. The relevance of testimony for knowledge cannot be underrated as many philosophers have noticed. Jenifer Lackey notes:

> "Almost everything we know depends in some way on testimony. Without the ability to learn from others, it would be virtually impossible for any individual person to know much beyond what has come within the scope of her immediate perceptual environment. The fruits of science, history, geography – all of these would be beyond our grasp, as would much of what we know about ourselves" (Lackey 2007).

Elizabeth Fricker has stressed that the role of testimony has even increased in our contemporary society and linked this to the increasing relevance of distributed labor. She argues that: "We citizens of the 21st century live in a world where division of epistemic labor rules. Most of what we know we learned from the spoken or written word of others, and we depend in endless practical ways on the technological fruits of the dispersed knowledge of others—of which we often know almost nothing—in virtually every moment of our lives" (Fricker 2006a: 592).

---

as sheer transmission should be overcome, the distribution of information is nonetheless a crucial socio-epistemic function of social media traditional models amongst others. Hence, classical analyses of testimony might shed some light on these processes of information transmission while they are not suited to analyze the role social media plays for knowledge creation.

Given this pervasiveness of knowledge received via testimony, one might assume that analyzing the epistemology of testimony has been a major topic in philosophy ever since. However, quite to the contrary, testimony has received quite little attention by philosophers until quite recently.

Epistemic Individualism in Western Philosophy

Many authors attribute the long-lasting neglect and depreciation of testimony to a pervasive *epistemic individualism* in Western philosophy. Some trace the roots of this individualism all the way back to Plato or Aristotle, while others stress the relevance of Descartes' individualistic rationalism as developed in *Discourse on the Method* (cf. Coady 1992: 4ff; Adler 2006). Yet, the most cited statement concerning the inferior epistemic status of testimony comes from Locke, who writes:

> "I hope it will not be thought arrogance to say, that perhaps we should make greater progress in the discovery of rational and contemplative knowledge if we sought it in the fountain, in the consideration of things themselves, and made use rather of our own thoughts than other men's to find it: for, I think, we may as rationally hope to see with other men's eyes as to know by other men's understanding... The floating of other men's opinions in our brains makes us not one jot the more knowing, though they happen to be true. What in them was science is in us but opiniatrety." (Locke 1961: 58, quoted from Coady 1992: 14).

This quote can be read as a direct request to use one's own cognitive capacities, one's own perception and inference in our attempts to know instead of relying on others. For Locke, perception, inference and memory were considered primary sources of knowledge, which might be deceivable but in principle reliable, knowledge observed via testimony seemed much more fallible. What we do not acquire through our own observation, through own inference or memory, cannot be considered knowledge. The problem with such a strong notion of knowledge is that most of what we would consider to be knowledge, the most basic examples of what we think we know, we actually would not know anymore: our date of birth, the size of Germany, the length of the equator, that Beijing exists if you have not been there, etc. Moreover, even *scientific knowledge*, by many considered to be the prime example of knowledge, would not be knowledge; since the majority of science is nowadays conducted in teams, scientists themselves often have to rely on the testimony of their peers (Hardwig 1991). Reducing knowledge to what we directly perceive or have inferred with our on-board cognitive resources would diminish enormously the body of what we consider to be knowledge.

So this is the dilemma of the epistemology of testimony: on the one hand, the majority of what we know, we know through others. As compared to the amount of knowledge that we receive via perception or inference, the vast majority of our knowledge is acquired from testimony. On the other hand, there is as of yet no consensus upon whether the knowledge acquired via testimony, through the words of others, even counts as knowledge.

The Rediscovery of Testimony

Despite this long-lasting neglect, the topic of testimony has received quite some attention in recent years. Several anthologies, monographs, special issues and numerous articles have been devoted exclusively to the epistemological analysis of testimony (e.g. Lackey 2007, Lackey and Sosa 2006, Coady 1992, Pritchard 2004, Lipton 1998, Fricker and Cooper 1987). Moreover, almost all authors portrayed in the previous chapter, have developed their own accounts of testimony or at least reference the topic (e.g. Longino 2002c, Kusch 2002, Kusch 2009c, Goldman 2003).

Given this newly inspired interest in the topic, I cannot do justice to the epistemology of testimony in this small section. Moreover, adding something new to this sophisticated debate in a just few words is out of the question. However, this is not the intention of this section. I do not aim at joining the highly specific and, to my mind, overly abstract debate around the epistemological problems of testimony in contemporary analytic philosophy. In lieu thereof, I sketch out some of the major issues and questions around testimony that are of relevance for the applied questions of this thesis, such as how people collectively construct knowledge and how these processes might be improved.

Two Major Debates: Inferentialism and Reductionism

There have been two central debates in contemporary epistemological accounts of testimony. One debate concerns the question of whether testimony is a basic source of knowledge with the same *status* as perception, memory or inference or whether its status is lower thereby requiring justification from more fundamental sources of knowledge. This debate concerns the possibility of a *global justification* of testimony. The second topic - about *inferentialism* – concerns the question of whether adoption of beliefs received via testimony is direct or inferential, meaning that it also involves an assessment of the honesty and competency of the testifier.

In the debates around the *status of testimony as a source of knowledge*, the two poles of argumentation are *reductionism* and *anti-reductionism*. Fricker summarizes the reductionist position by stating that "[r]eductionists about testimony hold that, if testimony is to be vindicated as a source not merely of belief, but of knowledge, our epistemic right to believe what others tell us must be exhibitable as grounded in other epistemic resources and principles-perception, memory and inference -- which are regarded by them as both more fundamental, and less problematic" (Fricker 1995a: 394).

The reductionist position is often traced back to David Hume's *Enquiry Concerning Human Understanding*, although the assumption that Hume is a reductionist is not undisputed (e.g. Faulkner 1998 & Welbourne 2002, cited from Kusch 2002: 31). Hume acknowledges the ubiquity of testimony and its relevance for knowledge by stating that "[...] there is no species of reasoning more common, more useful, and even necessary to human life, than that which is derived from the testimony of men and the reports of eye-witnesses and spectators" (Hume 1957: 88, quoted from Coady 1992: 7). Yet, Hume also asserts that we trust testimony only because we have repeatedly experienced – *via perception* - correspondences between testimony and reality. Hence, in those cases when other sources of knowledge are not at our disposal, we can trust testimony only *via induction*. The opposite position is often attributed to Thomas Reid's *principle of credulity*, according to which we have [...] "a disposition to confide in the veracity of others, and to believe what they tell us[...]", because God intended us to be "social creatures" (Reid 1983: 94f, quoted by Adler 2006).

The debate around reductionism versus anti-reductionism concerns the question of whether testimony has to be justified by other, "more basic" sources of knowledge, namely perception, inference or memory, or whether it is of the same status and therefore needs no further justification. Jonathan Adler summarized the difference between Hume's and Reid's view on testimony as follows: "Reid's position is that any assertion is creditworthy until shown otherwise; whereas Hume implies that specific evidence for its reliability is needed" (Adler 2006).

One major proponent of an *anti-reductionist position* today is C. A. Coady, who has written a seminal book on testimony (Coady 1992). He argues against the possibility of a testimony-free justification of testimony. While *global reductionism* is hardly defended anymore, a variety of more limited versions of testimony exist, such as Peter Lipton's concept of *rule-reductionism* (Lipton 1998), Fricker's *minimal or local reductionism* (Fricker and Cooper 1987; Fricker 2006a, as well as Goldman's *Bayesian justification of testimony*, (Goldman 2003) which has been depicted in some detail in the previous chapter).

The second main debate, concerning *inferentialism*, concerns the question of whether an assessment of the testifier's trustworthiness and competency are necessary for the rational acceptance of testimony. As expected, Coady is a proponent of a non-inferentialist position, arguing that acquiring knowledge through testimony is as direct as acquiring knowledge via testimony (Coady 1992). Since both perception and testimony are fallible yet reliable, he argues, no additional assessment of the competency and honesty of the speaker is principally needed. Elizabeth Fricker is a proponent of inferentialism. In a critical note on Coady's book on testimony (Coady 1992), she writes: "[...] it seems obvious that we must have a default position of trust in what others tell us - that human personal relations would be unimaginable without this; but consider others, and it seems equally obvious that our attitude to others must be critical and skeptical, that we must and do weigh the balance of probabilities against what they say being true, on the one hand, and against their being mistaken or insincere on the other" (Fricker 1995a: 406f). She concludes that we therefore have an epistemic duty to monitor the testifier for signs of insincerity or incompetence.

Testimony and Trust

While much of the debate portrayed above centers around the question of how and to which extent epistemic agents have to empirically test either an epistemic claim or the testifier in order to rationally accept testimonial evidence, there are alternatives to these *evidential* theories of knowledge (Faulkner 2007b: 877). Paul Faulkner, as do many other authors (Welbourne 1986, Hardwig 1985; Hardwig 1991), argues that the acceptance of testimony rests fundamentally on the question of whether the hearer *trusts* the speaker (Faulkner 2007b, Faulkner 2007a). In opposition to inferentialist accounts, Faulker insists that "[i]n trusting a speaker we adopt a credulous attitude, and this attitude is basic: it cannot be reduced to the belief that the speaker is trustworthy or reliable (Faulkner 2007a: 305) .

To my mind, many analytic accounts of the epistemology of testimony, such as those depicted above, describe "core cases" of testimony that are characterized by highly unrealistic situations. Adler (Adler 2006), for instance, delineates a model of testimony that is based upon several assumptions, including that testimony consists in a *brief assertion*, that the *hearer cannot attribute authority to the testifier* and that the testifier is a *stranger* to the hearer. Clearly, this might be a valid thought experiment to reach some general account of testimony. However it fails to address the question of whether such an abstract account has any relevance to real socio-epistemic practices.

As Faulkner has noted, such a view "[...] ignores the fact that the testimonial relationship is often embedded in an established relationship, or can initiate a presumption of relationship" (Faulkner 2007b: 876). Testimony most of the time is not a one-shot game. Rather, giving and receiving testimony involves the development of trust relationships between testifier and recipient. Trust is the *opposite* of relying on empirical evidence, because it implies dependence and the risk of being betrayed or let down (Faulkner 2007b, Mcleod 2006, Baier 1986).

Trust within Science

One of the authors active in the analysis of the role of trust in science is John Hardwig, whose work was introduced in the previous chapter. In his seminal paper, "The role of trust in knowledge", Hardwig assesses the function of trust for knowledge creation in science. He argues that in taking a closer look at scientific practice, we have to rebut the antithetical conceptualization of trust and knowledge still prevalent in epistemology that states "[w]e can not know by trusting in the opinion of others; we may have to trust those opinions when we do not know" (Hardwig 1991: 693). He notes that in epistemology as well as in philosophy of science, it has been nearly universally assumed that knowledge rests on evidence and not on trust. If knowledge were to rest on trust, it would have to be partly blind, which seems to be an unacceptable premise. However, Hardwig argues that analyzing current scientific practice even in the most renowned disciplines casts a new light on the relationship between knowledge and trust. "Modern knowers", he states, "[…] cannot be independent and self-reliant, not even in their own fields of specialization" (Hardwig 1991: 693).

His analysis departs from the observation that the majority of research is nowadays conducted in teams and he presents two exemplary case studies of major scientific achievements in physics and mathematics to support his claims. Cooperation in science is supposedly needed to overcome *time restrictions* on the one hand and to handle the rising *specialization* in science on the other. Due to this high specialization, scientists do not only lack the *time* to perform every subtask of their research on their own, but mostly they also lack the necessary *expertise* in the respective area of research. As a consequence, in scientific co-operations *scientists have to trust the competency and the honesty of their colleagues*. Moreover they have to rely on what Hardwig calls the "adequate epistemic self-assessment" of their peers, i.e. their ability to realistically assess their own competencies and levels of expertise in the areas of concern (Hardwig 1991: 700).

Thus, in order to successfully operate in science, scientists need to assess their colleagues not only *epistemically* but also *morally*. As a consequence, Hardwig emphasizes the necessity to consider and combine epistemology and ethics when reasoning about knowledge. To put it in a nutshell, Hardwig (Hardwig 1991) argues that trust is even more fundamental epistemologically than evidence like empirical data or logical argumentation because one needs to trust these pieces of evidence and their providers to actually use them at all. Thus, the trustworthiness of members of epistemic communities is fundamental to all scientific endeavors and represents the groundwork of (scientific) knowledge creation.

Trusting the Gentlemen: Steven Shapin's "A Social History of Truth"

One of the most interesting *historical* accounts of the role of testimony and trust for scientific knowledge was delivered by Steven Shapin. In his seminal book "A Social History of Truth," Shapin depicts the crucial role of gentlemen's testimony for the development of English experimental philosophy as a predecessor of experimental science (Shapin 1994). He shows how the development of this epistemic practice in the Seventeenth century fundamentally rests on the gentlemanly identities of its key advocates, such as Robert Boyle.

Shapin's premise is that "[k]nowledge is a collective good. In securing our knowledge we rely upon others, and we cannot dispense with that reliance. That means that the relations which we have and hold our knowledge have a moral character, and the word I use to indicate that moral relation is *trust*" (Shapin 1994: XXV). Shapin argues that our knowledge about the world depends on knowledge about other people and a crucial epistemic task therefore is to identify trustworthy people: Whom should I turn to in order to know? Whose testimony should I trust? In 17^{th} century England as much as in contemporary science, reliable witnesses are needed crucially to testify the results obtained by experiments. Hardwig gives a straightforward answer to the questions above (Hardwig 1991): trust people who are competent, honest, and who can assess their competency on the issue at hand.

But how should one identify those honest and competent testifiers? One good rule of a thumb would be to find someone who is *disinterested,* someone who is *free.* And in 17^{th} century England those who were considered to be free were the gentlemen, in particular because their economic position allowed them not to work. Thus, economic freedom was considered to be a prerequisite for moral freedom. It was assumed that "[g]entlemen were truth-tellers because nothing could work upon them that would induce them to be otherwise" (Shapin 1994: 84). Shapin argues that "[...] it was the *disinterestedness* of the English gentleman's situation that was most importantly identified as the basis of his truth-telling" (Shapin 1994: 83).

While some philosophers deplore Shapin's relativist notion of truth, (Lipton 1998: 4), or even more harshly criticize his work and consider it to be "social doxology" (Goldman 2003: 7), Shapin's account has deservedly received a lot of attention and praise. Feminist epistemologists and philosophers of science especially have built upon Shapins's work to analyze and stress the impact of gender and hierarchical social relations on the ascription of epistemic authority to knowers (cf. Anderson 2009, Alcoff 2001, Daukas 2006; Fricker 1998; Fricker 2006b; Fricker 2007). Such an extrapolation is clearly valid given that in 17^{th} century England women were deemed unreliable witnesses whose unreliable truthfulness was a result of their "constrained circumstances" and their economic dependence on men (Shapin 1994: 86ff).

Gentlemen were not only trusted because of their supposed moral and financial freedom. They also trusted one another because "they belonged to the same club" as Lipton notes (Lipton 1998: 12). Thus, at least two critical issues are related to the epistemology of testimony and trust: First, to what extent is the social identity of the testifier relevant for the attribution of epistemic trustworthiness and the evaluation of the testifier's epistemic claims (e.g. Alcoff 2001, Fricker 2007)? Second, to what extent does membership in the same community matter to the acceptance of testimony (e.g. Kusch 2002)? I return to both issues at various instances in the remaining chapters of this thesis as they are also fundamental for my analyses of epistemic social software.

Testimony and Social Epistemology

That there is a relationship between testimony and social epistemology should have become obvious by now. For the longest time in the history of philosophy, *addressing the social dimensions of knowledge simply meant analyzing the epistemology of testimony*. The question of how to treat testimonial evidence, knowledge reported from others, was the only way in which knowledge was understood as being social. To this extent then, the epistemology of testimony is a major philosophical predecessor of social epistemology. This view is supported by the fact that almost all social epistemologies portrayed in the last chapter have proposed their own accounts of testimony or at least addressed the issue.

Moreover, many social epistemologists still adhere to a very narrow conception of testimony as transmission of knowledge, which is also characteristic of the majority of analytic treatments of testimony described earlier. Alvin Goldman, for instance, has proposed a Bayesian account of testimony, which is clearly based on the model of testimony in a legal setting. Goldman proposes a simple sender-receiver model to transfer knowledge from one person to another. In this process, veritistic improvement can only be achieved by using communication to spread *existing knowledge* across a community; the generation of new knowledge via testimony is not considered. Opposing such a view of testimony as simple signal transmission, Martin Kusch argues that testimony is almost always generative of new knowledge. Criticizing social epistemologists for adhering too closely to a model of testimony as mere knowledge transmission, Kusch instead has proposed his own model of performative testimony to account for the creative and generative aspect of testimony understood as “learning from communication”(Kusch 2002: 18). While I agree with Kusch that testimony should be understood more broadly, such narrow conceptions are nonetheless of some interest for socio-epistemological analyses of one important aspect of epistemic social software: the distribution and spread of information and existing knowledge over multiple epistemic agents. Hence, insights derived from narrow conceptions of testimony may be useful to understand certain functions of social software (e.g. the distribution of information), yet they may be completely unsuited or even misleading if applied to other issues (e.g. the creation of knowledge in socio-technical epistemic systems).

Instead of rehearsing the arguments from the previous chapter, I conclude this section with two important epistemological views on the debate concerning the status of testimonial knowledge in comparison to knowledge from other epistemic sources: Martin Kusch's notion of *epistemic interdependence* and Helen Longino's disentanglement of *knowledge as knowledge-productive practices* and *knowledge as a state.*

Testimony and Epistemic Interdependence

Jonathan Adler notes that "[t]estimony depends upon other fundamental sources of epistemic warrant like perception or memory, but not conversely" (Adler 2006). Is this position plausible? Perception, to take one of the allegedly direct sources of knowledge, is a process of filtering information out of available information. Anyone whose mind has been occupied by something for quite a while might have experienced one the following situations – or a variation of it: seeing the initials of a loved one on many license plates, pregnant women walking all around the neighborhood, no-smoking signs on every corner, etc. In philosophy of science, similar phenomena are accounted for under the label '*theory ladenness of observation'*.

Clearly, perception, the extracting of information, is not random, but bound to memory, to inference and also to testimony. The dependence of memory on perception has been demonstrated over and over again in cognitive psychology. Elisabeth Loftus, in particular, has devoted much of her career to the so-called *mis-information effect*, which in fact is nothing beyond and a clear indicator of epistemic interdependence Loftus and Palmer 1974; Loftus and Hoffman 1989). Additionally, the relevance of *language* for all epistemic sources cannot be questioned, even if the extent to which testimony, perception, inference and memory are influenced by language may differ. At the very least, one has to acknowledge that a major part of reasoning depends on categories, which are socially constructed, i.e. based on linguistic conventions. If one takes this admittedly brief reference to empirical results seriously, then perception, memory and inference cannot be considered more fundamental than testimony. Those cases show that, although degree may differ, perception, memory and inference are also *socially shaped*. From the authors who are portrayed in this chapter, it has been Martin Kusch in particular, who has stressed the *epistemic interdependence* of all epistemic sources. Kusch argues that not only are "human cognizers [...] highly gregarious and deeply interdependent" (Kusch 2009c 20), but the sources of knowledge themselves are deeply intertwined (Kusch 2002).

If one adopts such an account of *epistemic interdependence*, as I do on *naturalist grounds*, we must reject reductionist claims stating that testimony depends upon more fundamental sources of epistemic warrant, such as perception or memory, while those sources do not depend on testimony (or other sources more generally).

Testimony: Epistemic Practices versus Knowledge as a State

In the previous chapter I introduced Hardwig's puzzling question of *who knows* when knowledge has been produced collectively (Hardwig 1991). In referring to an example where ninety-nine scientists collectively produced a paper on charm particles he asks who knows. If we insist that knowers must be able to provide evidence for all aspects necessary for the creation of knowledge and that knowers must be individuals, then no-one knows. While Martin Kusch (Kusch 2002) argues that the solution to this dilemma lies in acknowledging the community as the knower, Helen Longino proposes a different solution based on her three modalities of knowledge, *knowledge as content, knowledge as knowing* and *knowledge as knowledge production* (Longino 2002c).

Longino argues that differentiating between her three different modes of knowledge can clear much of the confusion about the epistemological status of testimony. As noted before, the reductionism debate around testimony concerns the question of whether knowledge obtained via testimony has a different, mostly lower status than knowledge received from other epistemic sources like perception, reasoning or memory. Longino argues that this debasement of testimonial knowledge has its roots in an extended individualism and a lack of consideration of the different modalities of knowledge - more precisely, in a conflation of *knowledge as knowledge-production* and *knowledge as the state of knowing*. Knowledge does not lose any value in transmission, she claims, and to know something, you do not have to have produced this knowledge yourself.

Referring to John Hardwig's article (Hardwig 1985) on the role of trust in knowledge, Longino argues that if it were the case that knowledge lost its quality in transmission, most contemporary practices of science could not be considered knowledge producing.

Following on this, if science, as the perceived stronghold of knowledge, is no longer considered emblematic of it any longer, then the question of what remains to be called knowledge arises. In fact, without such non-dissipative transmission of knowledge, not only would contemporary science be impossible, but the institutions of education and journalism would not make sense either!

One of the open questions in Hardwig's article is the question of who possesses knowledge that is collectively produced, i.e. who is the knower? Longino argues convincingly that once we separate the process of knowledge production from the state of knowing, this problem disappears. The community is the producer of knowledge (in Hardwig's case, the community of mathematicians and physicians Hardwig 1991). However, the moment this knowledge has been produced by the community, it can be obtained and possessed by anyone.

6.2.2 Distribution of Cognitive Labor

Another central socio-epistemic mechanism that has been extensively addressed in social epistemology is the distribution of cognitive labor. How should effort be optimally distributed over multiple epistemic agents? In principle, this topic can be addressed at the level of scientific communities or of smaller entities such as research teams. Within social epistemology, work using the label "distribution of cognitive labor" has focused primarily at the macroscopic level of research communities and has included the question of how resources and funding should be optimally distributed within a scientific community (Solomon 2001, Kitcher 1990). However, in the second part of this thesis I argue that the process of distributing labor over multiple epistemic agents is a fundamental socio-epistemic mechanism that operates on various levels. It is a major constituent of my socio-epistemological model.

Almost all social epistemologists address the distribution of epistemic or cognitive labor more or less explicitly.[100] One of the most explicit and formal treatments has been provided by Philip Kitcher in the last chapter of his book "The Advancements of Science" (Kitcher 1993). I have decided to portray his approach in some detail here for

[100] I prefer the term epistemic labor over cognitive labor, because it includes epistemic practices that would not fall under in the category of cognitive practices narrowly conceived as taking place within one's mind. Instead preparation of materials, the interaction with technological artifacts, the usage of mathematical algorithms, etc. would all be considered epistemic labor relevant for the creation of knowledge.

three reasons. First, I consider Kitcher's to be one of the most in-depth analyses on the topic. Second, his account is the one that is most frequently referred to within socio-epistemological discourse and influenced other treatments of the topic within social epistemology (e.g., Miriam Solomon's work). Third and most importantly, Kitcher distinguishes two central aspects of the organization of cognitive labor in science that are also important for of my own socio-epistemological model. The first one concerns scientists' relations to one another, including questions of trust and authority. The second aspect concerns the distribution of cognitive effort within scientific communities. I return to both questions in the third part of this thesis. In this section, I focus on Kitcher's work on distributed cognitive effort only.

Distribution of Cognitive Labor in Science

Distribution of cognitive labor is one of the two main socio-epistemological topics that Kitcher addresses in his chapter on "The Organization of Cognitive Labor" (Kitcher 1993). This section of his work refers to the question of how effort, resources and funding should be optimally distributed within a scientific community. His guiding questions include what amount of and division of effort is desirable within a scientific community and how diversity in scientific communities should be maintained. Kitcher proposes a *normative-prescriptive* model for the distribution of cognitive labor in science based on formal analyses, instead of an empirical account of how scientific effort actually *is* distributed in different scientific fields. This normative-prescriptive orientation is notable in the last sentence of a paper on the division of cognitive labor: "How do we best design social institutions for the advancement of learning? The philosophers have ignored the social structure of science. The point, however, is to change it."(Kitcher 1990: 22).

Kitcher's analyses of the distribution of cognitive labor are *socio*-epistemological with two respects. First, he describes an epistemically beneficial structure of a *scientific community*, i.e., the object of analysis and recommendation is not aimed at the individual scientist, but rather at the scientific community as a whole. Secondly, by highlighting the benefits of *epistemically sullied agents* Kitcher allows *non-epistemic motives* to play a *positive* role in the quest for knowledge. In continuing this section, I briefly summarize Kitcher's analyses, confining myself to a depiction of the general idea behind this model while omitting many details, including Kitcher's formalizations and his comments on the constraints of his model.

According to Kitcher, the value of an epistemic practice is judged in terms of the rate of progressive change. In turn, progress is judged in terms of the *attainment of significant truth (*Kitcher 1993: 118ff), a focus which Kitcher shares with Alvin Goldman (Goldman 2003). The question then arises: how should cognitive labor be distributed to enable the largest possible gain in significant truth?

Imagine a situation in which a scientist can choose between three different methods for approaching a certain problem. Let's assume further that one method – method A - is considered to be much more likely to lead to results than the other two. These other two approaches, B and C, are not completely implausible, but they are less promising than method A. If the scientist is to decide rationally, then for him it would be best to choose method A. But if this is individually the most rational decision and all scientists act rationally, then this would be *detrimental* to science. Why? Because from such an individual point of view, the optimal distribution of research effort would lead to a situation in which all scientists use them same method – method A. And this would lead to a *premature consensus method A*, although method B or C may eventually lead to the desired results. Whenever such a less promising method ends up being the one that solves a problem, premature consensus on the most promising method would have been detrimental for science. No one could have found out that the less promising method would have been successful, because no one had used it. But if a situation in which all scientists decide rationally leads to results which are detrimental for science, this means that *individual rationality and social rationality can contradict each other.* What is best for an individual scientist may not be what is best for the community as a whole.

To avoid the problem of premature consensus a plurality of methods and theories should be supported and cognitive diversity should be encouraged (Kitcher 1993: 344). More specifically, from the perspective of the community as a whole, the best thing is to *allocate scientists to each method in accordance with the relative plausibility of those alternatives.* How could such an allocation of scientists to tasks be achieved? How can we make sure that less promising approaches are also pursued, given that for each individual it is most rational to pursue the approach with the highest probability of being successful?

One could either hope for some very altruistic scientists who pursue the less promising approaches solely for the collective good of the scientific community. Or, less optimistically, one could imagine a dictator, or a *philosopher-monarch* to use Kitcher's expression (Kitcher 1993: 334) who has absolute control over the research activities in his empire. Given that he has an "[...] unerring eye for detecting the objective merits of theories" (Kitcher 1990: 8), he should distribute the scientific effort proportional to this merit. If method A is considered to have a probability of 0.6 to lead to true results, method B 0.3 and method 0.1, then our dictator should allocate 60% of the scientists to work on method A, 30% on method B, and 10% should on method C. Unfortunately – or luckily, depending on the point of view – there is no such science dictator in Western societies. Moreover, although science funding certainly has effects on the amount of scientific effort in different domains, allocation clearly is not that perfect. Furthermore, while science policy can *deny* funding for certain research programs, it cannot force scientists to pursue certain strategies or topics against their will beyond providing incentives for certain strategies or launching calls for specific areas of research.

However, maybe neither excessive altruism nor a science dictator is needed actually to solve the dilemma of contradicting individual and social rationality. According to Philip Kitcher, certain characteristics of scientists solve this dilemma without top-down allocation. The solution lies in *epistemically sullied communities (*Kitcher 1993: 345). Epistemically sullied communities consist of *epistemically sullied agents* who are not only motivated by their eagerness to find truth, but are also credit-driven. These agents strive for recognition from their peers, for prestige, for being first. The characterization of scientists as epistemically sullied agents appears descriptively quite accurate. Being the first person to discover something or to obtain significant results appears to be a crucial factor for many scientists, especially in the natural sciences.

If a scientist uses the same method as many other scientists, the probability that he will be the *first* to make a new discovery is quite low. On the other hand, even if his method is fairly unlikely to succeed given the available evidence, if it actually is correct and the scientist has been the only one who pursued it, he will earn all the merit. Therefore, a social factor mostly considered to be detrimental to science, namely the scientists' grubby quest for personal glory and reputation, is turned into a necessary condition for an optimal distribution of research efforts. Such Hobbesian communities comprised of epistemically sullied agents tend to work much better than communal approaches that are either purely rational or even altruistic (Kitcher 1990).

Kitcher concludes: "The very factors that are frequently thought of as interfering with the rational pursuit of science—the thirst for fame and fortune, for example—might actually play a constructive role in our community epistemic projects, enabling us, as a group, to do far better than we would have done had we behaved like independent epistemically rational individuals. Or, to draw the moral a bit differently, social institutions within science might take advantage of our personal foibles to channel our efforts toward community goals rather than toward the epistemic ends that we might set for ourselves as individuals."(Kitcher 1990: 16).

Division of Cognitive Labor Beyond Science

Although Kitcher's approach is developed in the context of science, questions of how to distribute cognitive labor are clearly also of tremendous relevance for other epistemic environments and practices. Consider different epistemic communities on the Web, such as the open source movement, the Wikipedia community, or an abundance of other epistemic communities that Yochai Benkler, one of the most influential thinkers analyzing the potential of the Web to transform modes of production, subsumes under the label of *commons-based peer production* on the Web (Benkler 2002).[101] For those projects, an enormous number of people need to be coordinated – or rather need to coordinate themselves and assume different tasks. Clearly, in most cases, there is even less formal coordination than there is in science with official science funding. Moreover, the modes of coordination and allocation typically differ, some might be more top-down, others completely bottom-up. Nonetheless, all these projects, many of which serve explicitly epistemic purposes, have to deal with the issue of how to organize and distribute cognitive labor.

A second aspect that has been stressed by Kitcher, and which also is of significant relevance for communities on the Web, concerns motivational issues. Some community members may be motivated by the reputation they can gain from their peers, some draw their inspiration from the joy of taking part in an enterprise centered on providing free knowledge, and others might take part in open source to learn some skills for their job portfolio in hopes of getting a job offer. Whatever their motives might be, the clear

[101] I will portray Benkler's work in more detail in Chapter 8.

effect is that an incredible number of people work for free by providing knowledge, information, software or other goods for themselves and others.

In focusing only on the distribution of labor within scientific communities, I have suppressed several other socio-epistemological topics that Kitcher has raised. These questions include the following: To what extent, when and with which consequences should we defer to authorities? How are authorities actually evaluated? How should credit be attributed? When should we cooperate and when is competition a better mode to achieve superior epistemic results? How should individuals or the community as a whole respond best to innovation? Is it better to be an early adopter or are laggards on the safer side? How and then should consensus be reached or resolved? How and to what extent is cognitive diversity needed? The topic of consensus formation is addressed below, and, as noted before, the issues of trust and authority become crucial in the third part of this thesis.

6.2.3 Consensus Formation

Consensus formation is another crucial topic in social epistemology and again, almost all social epistemologists deal with it in their theories. Some consider it to be crucial for the advancement of science (Kitcher 1993), while others argue that consensus should be avoided as long as possible (e.g. Solomon 2001). While some theorists provide normative models of consensus formation as rational processes of closure (e.g. Lehrer and Wagner 1981), others stress the dangers of premature consensus and deliver criteria for the dissolution of consensus (Solomon 2001). Although the stance towards consensus and the aspects of it that are emphasized may differ, the topic cannot be avoided in any social epistemology. Accordingly, consensus will also be addressed in my own socio-epistemological model later. There, I argue that consensus is a mode of closing socio-epistemic processes. Before this, however, let's take a look at a very specific theory of consensus.

Although several social epistemologists address the topic of consensus formation, I have chosen to portray Keith Lehrer's model of rational consensus in some detail to introduce the topic. Given that there are numerous models of consensus formation within philosophy, I deliver the reasons for choosing Lehrer's approach as a showcase. First of all, Lehrer's model is probably one of the most well-known models of consensus. Secondly, and possibly for that reason, it is the model of consensus formation that is most often addressed within socio-epistemological discourse. Although Lehrer is rarely directly labeled a *social* epistemologist and rather known for his coherence theory of knowledge (Lehrer 1990b), his work is central within the socio-epistemological field of discourse. Besides the reception of his model of rational consensus, his investigation of the relationship between personal and social knowledge should be mentioned, given that it was part of the crucial 1987 *Synthese* special issues on social epistemology (Lehrer 1987). Thirdly, Lehrer introduces a central distinction between *content information* and *social information*. This distinction and the relevance of social information as evaluative information about the competency and honesty of one's peers are crucial for the development of my own socio-epistemological model. The relevance of Lehrer's model for my own model is twofold: on the one hand it delivers an example as well as a normative model for a mechanism of closing processes in which multiple epistemic agents are involved. On the other hand, it is a mechanism that exploits *reputation*, i.e., social evaluative information to decide upon topic matters.

Keith Lehrer's Model of Rational Consensus

Lehrer's theory of rational consensus describes *how consensus should be rationally achieved* in science and society more generally, and as such is a *normative-prescriptive* account of consensus formation (Lehrer and Wagner 1981). This point is of socio-epistemological interest for two reasons. First of all, it describes a *social mechanism for achieving epistemic results.* Second, it deals with the question of how *empirical information* should be combined with *social information*, understood as information about the competency and honesty of peers, to achieve better epistemic results.

Together with the mathematician Carl Wagner, Lehrer has developed a formal theory of consensus in science and society more generally. It is depicted in great length in their book "Rational Consensus in Science and Society" (Lehrer and Wagner 1981), which consists of a philosophical section written by Lehrer and a mathematical section authored by Wagner. In the philosophical part, Lehrer situates his model of rational consensus in relation to philosophical accounts of consensus in different contexts. He argues that the role of consensus has been extensively discussed in a variety of different context and lists the major contributors to these debates. Those topics and their main proponent include politics (Hobbes, Rousseau, Locke and Spinoza), social choice (Arrow, Rawls), justice (Rawls), epistemology, science, intuition and common sense

(Reid, Locke, Moore, Chisholm, Goodman), and language (Putnam) (Lehrer and Wagner 1981: 3ff). Given the fact that consensus obviously has played and continues to play a profound role not only for philosophical reflection, but also for real-life decisions, a *rational* way of reaching consensus should be of prime concern for philosophy. Therefore, Lehrer combines these philosophical insights with statistical models for amalgamating information, especially of how to aggregate individual probability assignments to reach a consensual verdict ((Lehrer and Wagner 1981:17).

Lehrer and Wagner's theory of rational consensus rests upon the employment of *consensual probabilities*, utilities and weights for rational decision making processes in science, society and the arts (Lehrer and Wagner 1981, Lehrer 2007[102]). Lehrer argues that for decision making processes to be rational, it is central that *all* salient evidence or empirical information available be used. This spectrum of available information - for instance, concerning disputes on scientific theories - should not be limited to experimental information, however, but should also include the *opinions experts have of their fellows*. Lehrer calls this second type of information *social information* (Lehrer 1990a). Accordingly, his theory of rational consensus is meant to describe a rational form of information aggregation encompassing total empirical information – i.e., social information concerning the reliability of experts as well as the experimental data - that members of a given group or community possess collectively. Lehrer and Wagner's theory of rational consensus is a socio-epistemological theory, because it describes a socio-rational mechanism that can be used for epistemic purposes. It is actually social in a dual sense. Not only is a *social mechanism* developed to reach consensus rationally, but it is also *social information*, i.e., collectively assessed information about the intellectual authority, information about the reputation of scientists, that is epistemically used.

To illustrate how his model of rational consensus works, Lehrer uses the so-called "expert dilemma", a frequently encountered situation in which evidence for answering a question is inconsistent, thereby leading different experts to recommend different options (Lehrer and Wagner 1981). If scientific dissent is prevailing, but suspension of judgment is not an option, how should the conflicting information be used to reach a

[102] Lehrer was the invited speaker of the 15th Vienna Circle Lecture on the 24th of July 2007 in Vienna. In my account of his views I do also rely on his lecture "Consensus in Art and Science" given on that occasion.

consensual conclusion? “Decide on the experts to decide on the issue” could be the motto of Lehrer’s approach. Recalling that Lehrer proposes that the inclusion of all relevant and available information would be most rational, it becomes more plausible why he turns to assessing the other experts on the issue at stake before returning to the question itself. Each expert in a certain community might be more or less reliable or competent with respect to the specific question at stake. Therefore, it would be best to include each expert’s answer weighted by his competency. And who could assess each expert’s competency better than her peers in her scientific community? Since they are the ones who are most likely able to assess her competence, they could also *collectively decide* how much intellectual authority to grant her. It would be even better, *if each peer’s verdict on the other members of the group be weighted by his or her own ascribed intellectual authority*, and by his or her own *reputation* as well. Thus, in a collective process, through a *social mechanism*, quantitative weights are calculated for each member of a scientific community to serve as a quantitative indicator of this person’s intellectual authority on the issue of interest. One could argue that this amounts to using quantified reputational cues as information to decide on factual matters. This recommendation to *use reputation as an epistemic proxy* rests upon the assumption that members of a group—for instance, members of a certain scientific community—“have opinions about the dependability, reliability and rationality of the other members of the group” (Lehrer and Wagner 1981: 19).

Lehrer & Wagner develop a quite complex mathematical model that describes an iterative and collective process to reach quantitative values for the reputation of each scientist (Lehrer and Wagner 1981). The basic idea, however, is quite simple. The first step in this model consists in each expert giving a weight to all other experts summarizing all his information about the other’s expertise and reliability concerning the issue at stake. In other words, he gives a quantitative indicator of what he considers to be the reputation of the scientist with respect to topic at hand. In a second step, the average reputation values for each scientist are calculated with a specific algorithm and then laid open. Then in the second round, each expert has to reassess the reputation value he has given to all other members of the community, i.e., she has the chance to revise his or her judgment taking into account the average weights which the other members of the community have given to their fellows. Similar to Delphi-studies in the social sciences, this process is ideally repeated until finally a consensual weight for each member of a community is achieved (Linestone and Turoff 2002).

The idea is that, if you are less secure about the reputation of a certain researcher, you might tend more towards the group average in your second vote. If you are very sure about the reputation of someone, however, you will not let yourself be influenced by this average. If everyone acts this way, that is considered to be most rational, then the consensus that is finally achieved is considered to be the most rational consensus. According to Lehrer "[t]he aggregation procedure is simply an iterated averaging procedure for obtaining a summary of the total information in the initial state. The summary consists of a set of weights in which one convergent weight is assigned to each member of the group" (Lehrer 1990a: 176). Crucially, once these consensual weights are achieved, they can be applied to answering the question of concern by weighting each member's vote on the issue with their consensual personal weight of reputation.

Rational Consensus Beyond Science

Lehrer and Wagner (Lehrer and Wagner 1981) deliver a formal model of how a community should use social information, i.e., information about the intellectual authority of other people, rationally for epistemic purposes. Their approach is socio-epistemic to the extent that it describes a mechanism in which not only a social process is considered to be rational at the same time. It also argues for the epistemic value of social information, i.e. judgments about others' reputation and the intellectual authority they should be granted with respect to a certain topic matter.

For the purpose of this thesis, this *epistemic use of reputation* is highly relevant. Moreover, taking into account the prevalence of all sorts of indices, which are meant to function as proxies for the reputation of researchers and even the quality of their work (e.g., h-index, etc.), taking a closer look at how these *algorithms* actually function might prove rewarding. Clearly, not all epistemic usage of reputational cues has to follow such a formal method. Quite on the contrary, ratings and other reputational tools might be used in a variety of different ways on the Web and our everyday life more generally. Nonetheless, Lehrer & Wagner's model delivers a clear example of the potential that reputation, particularly when understood as social information from an evaluative stance, can have for epistemic tasks (Lehrer and Wagner 1981).

That said, however, there are limits, shortcomings and dangers inherent in Lehrer and Wagner's model. The first problem concerns the concept of the communities. Although the community of scientists plays a decisive role in Lehrer's model, it remains underdetermined. The whole process of rational consensus formation is based on the ability of experts to evaluate their fellows. However, how these judgments are to be obtained in the first place remains open. It is obvious that for experts to judge the competency of their fellows, they must have had contact with them at least indirectly via reading their publications, etc. However, even though this interaction and communication within the communities is a necessary prerequisite for the whole process of obtaining consensual weights, it is not explicitly described or analyzed in Lehrer's approach. The only form of communication which is explicitly needed for the process of consensus attainment is the interchange of numbers, i.e., the communication of weights. The social or community aspect of knowledge may rest upon information about the fellow experts, but the process of how experts arrive at their weights, how they assess their peers intellectual authority is not addressed.

Of particular concern here is the relevance of various societal factors (institutional background, gender, nationality) in this process. Given the centrality of reputational weights for consensus formation in this model and the fact that it is a normative model, one would assume that the processes of attributing reputational values to others is under close scrutiny in Lehrer's model. However, the criteria upon which epistemic agents form their opinions on their peers are not addressed at all within Lehrer's approach.

This neglect of the social processes taking place within scientific communities, as well as the broader societal context in which these are embedded, leads me to a more fundamental critique of Lehrer's model of rational consensus. The use of reputational cues, as beneficial as it might be for reaching consensus, comes with serious *epistemic and ethical dangers*. Primarily feminist epistemologists as philosophers of science have pointed out the risk of *epistemic injustices and biases* that can occur when invalid reputational proxies are used to assess knowledge claims without critical reflection. (cf. for instance Fricker 2007, Scheman 2001, Alcoff 2001). Feminist theoreticians have also convincingly argued that such biases and the unfair attribution of intellectual authority are not only ethical but also cognitive and epistemic failures, and have therefore requested a *tempered equality of intellectual authority* (Longino 2002c:131).

To be sure, neither the impact or relevance of communicative processes within social groups or communities for knowledge nor the potential infiltration of societal injustices into this process is negated in this account. However, these processes are rendered invisible by being used as important but unstudied prerequisites or as interchangeable sources of information or domains of application, and are thus removed from the center of inquiry. At various points in the course of this thesis I return to the possibility of epistemic injustice that lurks in the corners of social mechanisms that rely heavily on the collective attribution of intellectual authority.

To conclude, the use of reputational cues to weight knowledge claims according to the perceived reputation of the knower comes with opportunities on the one hand and dangers on the other. Reputation can help to make a quick judgment concerning the expected quality of information and, in this role, reputational cues are of high relevance on the Web , which is characterized by its overabundance of information. Reputation can serve here as a filter which can result in new *classifications of information,* which would be *second-order knowledge* (Origgi and Simon 2010). On the other hand, there is always the danger that invalid proxies are used to judge the epistemic trustworthiness of a knower: whenever social criteria that are not related to someone's quality as a knower are assessed to judge one's quality as a knower such epistemic injustices may occur. Hence, although one should consider using reputational weights for epistemic purposes, such usage and the processes of attributing weights has to be monitored and re-assessed constantly. I return to this topic in more detail when outlining the difference between weighted and unweighted forms of epistemic sociality in the third part of this thesis.

6.3 Central Debates within Social Epistemology: Differentiating the Field

As noted in the previous chapter, social epistemology, even when it is narrowly conceived, is not a homogenous or unified field. Rather, there are numerous controversies concerning even the most basic features of social epistemology. In the following section, I outline some of the most central debates within social epistemology. Clearly, the topics and categories I have chosen are neither exhaustive nor selective. Nonetheless, I think that they enable a clearer picture of the field of discourse known as "social epistemology". My goal in the following sections is neither to close any of the debates nor to contribute substantially to them. Rather I use these conflicts to span and outline the field and situate my own approach therein.

6.3.1 Complementarity

One of the central debates within social epistemology concerns the relationship between social and individual epistemology. Many social epistemologists consider their social epistemologies – or social epistemology in general – to be complementary to individual epistemologies, as traditionally conceived. Given the background and training of many social epistemologists, such individual epistemologies are mostly rooted in analytic epistemology.

Alvin Goldman is one of the main proponents of such a complementary view that distinguishes individual and social epistemologies as two distinct endeavors. In proposing to divide epistemology into two branches he sees their respective duties as follows: "Individual epistemology would identify and evaluate psychological processes that occur within the epistemic subject. Social epistemology would identify and evaluate social processes by which epistemic subjects interact with other agents who exert causal influence on their beliefs."(Goldman 2006). Other social epistemologists argue along the same lines. Hilary Kornblith, who has proposed a "conservative approach to social epistemology" in Schmitt's anthology (Schmitt 1994b), also considers social epistemology to be an amendment to individual traditional (naturalistic) epistemology (Kornblith 1994). While acknowledging that "knowledge is a socially mediated phenomenon" (Kornblith 1994: 97), he favors an account of social epistemology, which "will leave important room for input from the study of group processes and institutions, but will leave the overall structure of our epistemological theories substantially unchanged" (Kornblith 1994: 94). He states that "since belief acquisition and retention must be seen as a product of both social and nonsocial factors [...] both kinds of factors will come in for investigation in the course of epistemic evaluation" (Kornblith 1994: 102).[103]

[103] Interestingly, when referring to the nonsocial factors, Kornblith refrains from talking about *knowledge*, but instead refers to *beliefs*. That *beliefs* may be nonsocial is indeed not denied from all social epistemologists who argue that knowledge is social in a more profound sense. Helen Longino (2002c) for instance argues that while beliefs can be held by individuals, *knowledge* depends on discursive interactions in communities and therefore cannot be conceived nonsocially.

Philip Kitcher, whose work on the distribution of cognitive labor I discussed previously, has advocated for a distinction between individual and social epistemology as two complementary programs. While "[...] individualistic epistemology concerns itself with those processes that promote an individual's attainment of true belief, so too social epistemology should be concerned with the organization of communities of knowers and with the processes that occur among knowers within such communities that promote both the collective and the individual acquisition of true belief "(Kitcher 1994: 113).

Other social epistemologists, most notably Martin Kusch but also Helen Longino, argue that such a complementary understanding of social epistemology as a mere addendum that leaves the individualistic foundations of analytic epistemology untouched is insufficient. Both authors consider knowledge to be social in a fundamental sense and hence propose social epistemologies that question the foundations of individualistic (analytic) epistemology. Kusch instead proposes a *diagnostic social epistemology,* which instead of remaining as deferential as Alvin Goldman and other proponents of the complementary view "tries to analyze, explain and criticize the foundations of classical epistemology" in social and political terms (Kusch to appear: 3). Kusch considers his *Communitarian Epistemology* to be an example of an analytic version of such an diagnostic social epistemology. It is based on the conviction that "a social isolate is unable to know anything" (Kusch to appear: 4)) and that instead knowledge is a social status that communities grant to individuals or groups (Kusch 2002).

Although Longino puts certain emphasis on the role of the individual, her concept of knowledge is thoroughly social as well. This becomes obvious in her statement that knowledge, as opposed to belief, is a social status – and not a psychological state (Longino 2002c: 204). When analyzing observation and reasoning as two central epistemic practices in science, she states that social mechanisms are "[…] what enables the transformation in assertability-status from'It seems to me that p' to 'P'" (Longino 2002c: 103), and concludes that "[...] the results of both reasoning and observation, then, are socially processed before incorporation into the body of ideas ratified for circulation and use, or are treated as having been so processed" (Longino 2002c: 106).

Concerning the issue of complementarity, I side with Martin Kusch and Helen Longino. Knowledge has to be understood as social in a more fundamental sense than complementary approaches convey. Both authors consider knowledge to be a *social* *status* and highlight the central role of communities for the generation of knowledge. However, Kusch and Longino use different antonyms for the term "social status", and by so doing refer to central aspects of the sociality of knowledge. Kusch uses the differentiation into *social, artificial and natural kind* to characterize the *extent* to which these different kinds of things depend on communities, and asserts that knowledge as a social kind ceases to exist once the communal performative is withdrawn (Kusch 2002: 165ff). Longino, alternately, stresses that knowledge depends on interaction between multiple epistemic agents by defining it as a *social status* as opposed to a *psychological state* (Longino 2002c: 204).

To be sure, arguing that knowledge is thoroughly social, that it is a social status and that there is no such thing as an isolated knower does not imply that an individual cannot engage in *epistemic practices* or that individuals cannot *believe, see, remember* or *infer* something. This argument does imply, however, that only through the community can belief turn into knowledge, that only through communal vetting will the results of individual – or collective – epistemic practices be accepted as knowledge. Longino's critique of knowledge as a psychological state leads to the next topic: the concepts of knowledge employed in different social epistemologies.

6.3.2 What's Knowledge and Who's to Decide? The Reception of Sociological Research

One of the central demarcation lines in social epistemology concerns the stance towards the definition of knowledge as justified true belief or some variant of it. Alvin Goldman has used this criterion to distinguish *classical and anti-classical social epistemologies* (Goldman 2006). Those social epistemologists that adhere to a notion of knowledge as justified true belief are classical social epistemologists (e.g. Goldman 2003, Kitcher 1993) and those that reject it are anti-classical social epistemologists (Barnes and Bloor 1982, Latour 1987, Fuller 1988). Goldman further divides the classical social epistemologists into those that emphasize the role of social factors for the acquisition of *true beliefs* and those that highlight the sociality of *justification and rationality*. The proponents of the anti-classical positions do - according to Goldman (Goldman 2006) - neither care about notions of truth nor of justification, but instead take for *knowledge what is believed*, i.e., they deny the existence of any criteria which could distinguish knowledge from belief or what is true from what is merely taken to be true. According to Goldman, such approaches should not be considered proper epistemologies and he proposes the term “social doxology” (Goldman 2003: 7) for these positions instead. In a later paper, he labels those approaches that do not accept knowledge as (justified) true belief or oppose a correspondence truth, *revisionist social epistemology,* but nevertheless maintains his verdict that those approaches should not be considered real epistemologies. Such undeserving approaches include quite a variety of *isms*, namely: “[p]ostmodernism, deconstructionism, social constructionism, relativism, and the social studies of science, including the ‘strong programme’ in sociology of science” (Goldman 2009b). In the development of his own, *Veristic Social Epistemology*, Goldman criticizes those anti-classical positions and labels their proponents “veriphobes” (Goldman 2003: 7).

Martin Kusch and Helen Longino have both criticized not only the reduction of knowledge to mere variants of justified true belief, but they have also outlined serious flaws in Goldman’s reception of the theories he dismisses. Kusch acknowledges that among the many approaches that Goldman dismisses out of hand (such as postmodernism, deconstructionism, social constructionism, relativism and social studies of science) only the social studies of science have been analyzed – although this analysis is considered inaccurate. While Kusch (Kusch to appear) clarifies Goldman’s misrepresentations of the positions of David Bloor, Barry Barnes and Harry Collins, Helen Longino (Longino 2002c: 44ff) shows similar flaws for the depiction of Bruno Latour’s (Latour 1987) as well as Steven Shapin’s and Simon Schaffer’s (Shapin and Schaffer 1985) work. Clearly, one cannot critically assess all other approaches that address the sociality knowledge. However, the moment one uses them as a background foil to develop one’s approach against them, one should take great care in portraying them correctly. Otherwise, one conducts one of the fallacies, Goldman rightly indentifies as detrimental to science and knowledge: the straw man fallacy (Goldman 2003: 153): misrepresenting others’ approaches to make them more easy to attack.

Concerning the stance towards the social studies of science, I side with Martin Kusch and Helen Longino’s positions. Kusch has revealed his debt to the sociologists of knowledge, most notably to the works of Barnes, Bloor, Collins, and Shapin (Kusch 2002: 4)). Helen Longino has been particularly receptive to the works of Bruno Latour and urges philosophers of science and social epistemologists to take science studies serious more generally (Longino 2002c). I agree that social epistemology would lose a central source of information if it does not consider sociological analyses relevant. Any consideration of how epistemic practices *should* be conducted has to understand first how they *are* conducted. This stance does not imply that sociological methods are superior to philosophical method, nor acts as a judgment about the relative worth of philosophical versus sociological, or theoretical versus empirical reasoning. Rather, both fields should be more receptive to one another. Empirical analyses deliver crucial insights for theoretical reasoning; they are important as a source or as a probing stone for normative considerations. Moreover, the social studies of science and technology have delivered invaluable insights for the analyses of socio-epistemic practices in general and socio-technical epistemic practices in particular, as I have shown in Part 1.

Longino's *Critical Contextual Empiricism* is also an attempt to build a bridge between philosophical and sociological analyses of science and knowledge by providing a framework that is "[…] responsive to the normative uses of the term 'knowledge' *and* to the social conditions in which scientific knowledge is produced." (Longino 2002c: 1). Remember that one of her primary goals lies in dissolving the *social-rational dichotomy,* which she holds responsible for the continuous misunderstandings between philosophers and sociologists of science.

To overcome this dichotomy, Longino has proposed a tripartite notion of knowledge that serves as a framework for my own analyses. Longino distinguishes *knowledge as cognitive agency* from *knowledge as knowledge-producing practices* and *knowledge as content*. Knowledge as cognitive agency, knowledge as the state of a person towards an object and the discussions around the formulization "S knows that p", refer to merely one modality of knowledge. Longino asserts that in each modality knowledge is social and rational at the same time; in each case "[t]he social is not a corrupting but a validating element in knowledge" (Longino 2002c: 122). Stressing the interdependence of cognitive agency, the plurality of knowledge and the contextuality of knowledge-producing practices, she proposes normative recommendations that take into account the actual knowledge-producing practices in contemporary science, thus consolidating sociologists' and philosophers' concerns.

Thus, to understand the *sociality of knowledge* comprehensively, this focus on *knowledge as cognitive agency* has to be amended at least by a comprehension of *knowledge as knowledge-producing practices* as well as *knowledge as content*.

6.3.3 The Utility of Truth

It should not come as a surprise that a concept as central for epistemology as "truth" has received a lot of attention in social epistemology. The previously portrayed debates around the notion of knowledge as justified true belief – or true belief only (Goldman 2003) – are a clear indicator of the relevance of this concept for socio-epistemological considerations. Nonetheless, the utility of the notion of truth for socio-epistemological analyses is less then clear. While truth and the truth-conduciveness of socio-epistemic practices loom particularly large in Goldman's *Veristic Social Epistemology*[104], other social epistemologists have assessed the merits and shortcomings of the concept of truth for the assessment of epistemic practices and products more critically.

With respect to science for instance, different social epistemologists have noted that truth might not be the best indicator of the quality of *scientific theories* (Solomon 2001; Longino 2002c). Miriam Solomon notes that the majority of theories in the history of science have been neither literally, partially nor approximately true (Solomon 2001). There may be some truth in scientific theories that have proven successful, but this truth can be recognized only in hindsight. The truth value of a scientific theory can be assessed only at a later stage than when it was first proposed. Even if knowledge is considered true at this later stage of appraisal, it is impossible to know whether or not it could be falsified in the future. It is this dependence on the historical reconstruction of truth, the "[p]ractise of assessing truth in the past theories from the perspective of present knowledge" (Solomon 2001: 33), that made Solomon choose the label "Whig" for her *Whig realism*.

Instead of stretching the concept of truth to allow for the assessment of scientific theories, Helen Longino has proposed the concept of *conformation* as an umbrella term for the relation between the content of knowledge and its intended object (Longino 2002c: 117ff). Truth is but one instantiation of conformation; others are fit, approximation, etc. I side here with Longino because there are several reasons why conformation may be an appropriate concept for this thesis. First of all, the concept of conformation can be applied to non-propositional content: diagrams, maps are not true

[104] Goldman adopts a correspondence theory of truth with some elements of deflationary approaches. The core of such a correspondence theory for Goldman is that "what makes sentences or propositions true are real-world truth makers" (Goldman 2003: 68).

or false - they are more or less useful for certain tasks.[105] The same is true for epistemic practices, they are also neither true nor false – they are only more or less useful for the achievement of epistemic goals. For his *Veristic Social Epistemology* Goldman has to introduce the concept of *instrumental veristic value* to retain the notion of truth for the assessment of epistemic practices (Goldman 2003: 87). Moreover, conformation, as opposed to truth, comes in degrees and it can be assessed for different dimensions depending on the purpose of epistemic tasks. All these characteristics make conformation an attractive concept for analyzing socio-epistemic practices and products on the Web.

6.3.4 The Relevance of Epistemic Communities: Contextualism and Relativism

Epistemic communities are a central concept in social epistemology. Every social epistemologist refers to groups of epistemic agents in some form or another. The minimal acknowledgement consists in agreeing that knowledge can be and often is created collaboratively in groups and that such collaborations may be epistemically beneficial, etc (e.g. Thagard 1997a). Other authors analyze the ontological and epistemological status of groups by asking what it means for a group to know, etc. (most notably Margaret Gilbert, e.g. Gilbert 1987; Gilbert 1989; Gilbert 1994; Gilbert 2004, but also Schmitt 1994c, Mathiesen 2005, Tollefsen 2007, Tuomela 2004). In most approaches it is acknowledged that knowledge can be created in groups, but it is usually posited that individuals can create knowledge as well.

Some authors go further than this, however, and insist that knowledge *depends* on epistemic communities in a more profound sense. Martin Kusch and Helen Longino, in particular, have put emphasis on the role of communities to grant knowledge as a social status, to provide the means for vetting content *as knowledge*. Without a community to declare something to be knowledge, there can be no knowledge. Individuals may have doxastic autonomy, they may not be constrained by the *beliefs* held in their community, but *knowledge* requires discursive interaction and *being a knower* requires *being recognized as a knower*. As Helen Longino asserts: "'S knows ...' attributes a status to S

[105] Of course maps can be false if they do not correspond to the terrain they intend to correspond to – this would be a case of non-fit, a lack of conformation. It is where reality bites back. Yet conformation as a term indicates that the evaluation of epistemic practices and tools depends on their intended purpose.

on the basis of S's actual engagement in or presumed capacity to engage in certain discursive interactions. Cognitive agents are interdependent, and their knowing is a social status, not a kind of psychological state" (Longino 2002c: 204).

But how fundamental is such a dependence of knowledge on communities and what are the consequences? Does the fact that only communities can grant knowledge imply that knowledge is only valid within specific epistemic communities? And what is the status of knowledge in epistemic communities other than the one where is was initially created? Can knowledge claims be detached from their contexts of creation? And how then can communities be wrong? These topics refer to the relationship between contextualism and relativism. Both, Martin Kusch and Helen Longino advocate for contextualist epistemologies, in which the existence of epistemic communities is a central prerequisite for knowledge. But the two authors differ profoundly in their answers to the questions posed above. While Helen Longino argues that her approach is contextualist, yet decidedly non-relativistic, Martin Kusch embraces a methodological relativism.

Helen Longino: Contextualism without Relativism?

Longino defines cognitive communities as "[…] any group bound by some set of common goals and shared public standards regulating critical (knowledge-productive) discourse and the stabilization of representations as knowledge" (Longino 2002c: 145)). While such communities are bound together by common goals and shared standards, these goals and standards differ between different communities. Even within contemporary Western science, which is Longino's frame of reference, different epistemic communities, even within scientific fields, have different goals and standards for evaluating epistemic practices and products. Controversies within scientific fields are an indicator of such epistemic pluralism.

Based on the acknowledgement of epistemic pluralism within science, Helen Longino advocates for *local epistemologies*. Having analyzed various case studies in biology, she argues that "[…] a plurality of adequate and epistemically acceptable explanations or theories can be generated by a variety of different factors in any situation of inquiry" (Longino 2002c: 184). These different factors include various *methodological and substantive assumptions* as well as premises about the *form of knowledge*. *Methodological assumptions* include decisions about appropriate methods of data generation (e.g. field work versus experimental studies), the kind of data (qualitative or quantitative) as well as assessments of reliability, etc. *Substantive assumptions* are assumptions about the world under study, such as the acceptance of the regulative function of genes. Finally, *forms of knowledge* are supposed to describe what is often called epistemic virtues, i.e., virtues such as simplicity or unification. However, all of these virtues, Longino argues, are not only based on substantive assumptions themselves, but are also hardly universal. The only *universal* standard she accepts is that of *empirical adequacy*, which has to guide every scientific endeavor. Any community of researchers has its very own assemblages of these different background assumptions, which form this community's "local epistemology". However, this local epistemology, as a "dynamic complex of beliefs, norms, goals, and practices" (Longino 2002c: 187), is bound to change when confronted with external or internal challenges. Such challenges might either be encounters with researchers having different background assumptions, changing goals, or data that seem to produce internal inconsistencies or tensions. And in preparation of what follows in the next chapters, I would like to argue that the introduction and use of new technologies, resp. the establishment of new socio-technical assemblages, is another source of such epistemic change.

Accounting for these background assumptions and their impact on research does not imply that these assumptions are necessarily detrimental or that they can or should be abolished. However, recognizing the situatedness of research and the partiality and plurality of knowledge should make knowledge claims more modest as well as appraisal of alternative views more palatable. Longino therefore concludes (Longino 2002c: 202): "The challenge posed by the plurality of contemporary science is not so much a metaphysical one but an educational one. How can the value of scientific research as a source of guidance for policy decisions be maintained in the face of the complexity of nature and the partiality and plurality of our knowledge of it?"

While I completely agree with Longino's analyses and her arguments concerning the plurality, partiality and provisionality of knowledge as well as the necessity for educating people to acknowledge these characteristics, I am not sure whether such an account can defend relativism in the end. Let me elucidate. Longino argues that the sociality of knowledge or the role of communities in knowledge production does not mean that the members of the community *share* concepts, goals or theories, but rather that they are *in interaction with each other* on these concepts, goals and theories. What is needed for this interaction to take place, however, are *shared* standards of criticism and evaluation.

These shared standards are Longino's weapon against relativism. But are they effective? Longino argues that "[s]hared standards permit diversity of beliefs, but unity in their methods of evaluation" (Longino 2002c: 148). But how do these standards come into play? And are they stable or do they change over time? Longino suggests that while standards are temporarily binding, they indeed change over time and differ between different communities. But if that is the case, how can they be said to defeat relativism?

If relativism is not meant to denote that anything goes, but rather that epistemic practices and knowledge can only be evaluated *relative to certain criteria*, to a specific set of standards, then Longino's approach is relativistic. And the argument that while a framework of reference might be chosen more or less freely, each choice comes with constraint is neither new nor to my mind unacceptable. Indeed, I think it is the only empirically adequate and consistent understanding of what contextualism may amount to. Similar ideas have been proposed by numerous theoreticians, such as Ian Hacking, Yehuda Elkana, Luwdik Fleck.

Yehuda Elkana (Elkana 1986) argues that in choosing a frame of reference *conventionality* comes into play and one has to be a *relativist with respect to possible frameworks*. Truth and logic are always formulated in accordance with this chosen frame of reference, a certain culture and language, and this contingency cannot be avoided. However, once a certain framework is set, criteria of truth and rationality can be set. Hence, Elkana considers himself to be a *relativist between frameworks, but a realist within a framework*. He labels this view historical relativism. As compared to stronger relativists, such as Mary Douglas, who has had a huge impact on the Edinburgh School, Elkana stresses the belief in reality and objective realism based on consensus. Along with Longino, he also stresses the necessity of critical dialogue as a prerequisite for the growth of knowledge.

Ian Hacking (Hacking 2002: 4), referring to Comte's and Foucault's notion of positivity, uses the concept of *truthhood* and *falsehood*, arguing that truth conditions are bound to *styles of reasoning*. Every *style of reasoning*, such as the statistical or the laboratory style, enables new conditions for truth, new ways of approaching the truth and new criteria for argumentation and demonstration. The question accordingly is not what is true and what is false, but rather *how* is something true? What are the conditions for something to be true?

Ludwik Fleck (Fleck 1980), whose concept of thought style ("Denkstil") forms one vantage point for Hacking's style of reasoning, introduces the notions of *active and passive connections* to account for the relationship between construction and cognition on the one hand and reality on the other. While active connections are contingent and depend on the prevailing thought style, the passive connections *impose themselves upon the observer*. They are "reality biting back" -- to use Longino's terminology (Longino 2002c: 119).

Thus, according to Elkana (Elkana 1986), Hacking (Hacking 2002), Fleck (Fleck 1980) and Longino (Longino 2002c), epistemic standards and conditions for truth are historically and locally contingent. Nonetheless, success is not arbitrary and construction is not just "making things up". As Longino concludes: "[s]uccess cannot just be a matter of the user wanting the theory to be correct [...]. If they are not, reality will eventually bite back. To suppose that success has nothing to do with the interests of the users is just as fruitless, however. The idiom of representation must be such as to enable successful interaction with that which is represented. The choice of idiom, and of the degrees to and respects in which it must fit the objects of representation, is a social choice, a matter of goals collectively endorsed in the community conducting inquiry. In this sense the community determines what will count as knowledge" (Longino 2002c: 119). But doesn't this form of contextualism also imply relativism understood as evaluation *relative to contextual criteria and local standards?* This would be exactly Martin Kusch's position, who is a declared relativist. He states that relativism as he uses the term does not mean that " [...] all beliefs or statements are only relatively true" (Kusch 2002: 269), but rather that truth and falsity depend on a communities interests, exemplars and goals.

Martin Kusch: Embracing Relativism

In Martin Kusch's *Communitarian Epistemology*, the community looms particularly large. Indeed, the community is the primary knower. There is no such thing as an isolated, individual knower - without communities, there can be no knowledge. But while Longino considers her theory of *Critical Contextual Empiricism* to be contextualist but non-relativist (Longino 2002c), Kusch complements his

Communitarian Epistemology with a “communitarian finitistic form of relativism concerning truth, reality, and objectivity”(Kusch 2002: 280). Although he assures his readers that it is possible to accept his positions on testimony and empirical belief without having to endorse his stance on relativism, he has continued to work on developing an account of methodological epistemological relativism more recently (Kusch 2009b, Kusch 2007).

Knowledge in Kusch’s account refers to a “bundle of entitlements and commitments” (Kusch 2002: 166), which link together those who know. Entitlements and commitments relate us to one another in our knowing. As Kusch states, “[i]n claiming to know something, we commit ourselves to being able to marshal evidence in support of the claim. We entitle our interlocutors to call on us to present this evidence, and – at least prima facie – we entitle ourselves to the status of someone who is worth consulting and following” (Kusch 2002: 166).

In his characterization of epistemic communities, Kusch combines Welbourne’s idea of *primitive communities of knowledge* (Welbourne 1986) with Robert B. Brandom’s notion of *entitlements and commitments* (Brandom 1994). Brandom uses this notion for his account of objectivity, arguing that the game of giving and asking for reasons forms the basis of all our social discursive practices. Thus, epistemic notions such as objectivity, knowledge, or truth can only be understood in relation to these discursive practices. Despite this indebtedness towards Brandom, there are crucial differences between his and Kusch’s conclusions concerning the consequences of community centrality for knowledge. More specifically, if it is the community that decides upon and attributes knowledge, how can the community be wrong? And does the fact that knowledge depends on communities implicate that knowledge is always only relative to this community?

Brandom rejects the idea that objectivity is based on intersubjectivity, because such a concept of objectivity based on consensus in a community would not allow for a community as a whole to be wrong about something. By contrast, he states, it is exactly this difference between what is *true for us* versus what is *true in fact,* which allows objectivity to come into play (Brandom 1994, as cited in Kusch 2002: 215). Thus, Brandom argues that I-thou relationships are more fundamental than I-we relationships, which would be the central term in communitarian epistemologies. If I-we relations were fundamental, there would be no possibility for a community to be wrong. However, it is exactly this possibility of error that needs to be accounted for if *truth* differs from *what is taken to be true*.

Kusch considers Brandom's argument for the primacy of I-thou-relationships to be unconvincing. Admitting that a community can be wrong, Kusch insists that "[…] the judgement according to which a whole community is, or was, wrong can only come from within another community (or a later time-slice of the same community). And such *judgement* will have to be based on the exemplars and the consensus in that other community" (Kusch 2002: 259). This implies that there is no view from above, that criteria of evaluation are always local and contingent. Indeed, Kusch acknowledges that "we cannot escape our own contingency as members of our own culture and its tradition" (Kusch 2002: 274). But as noted before, while Longino considers her approach to be contextualist but non-relativist because of the relevance of shared standards, Kusch embraces relativism by stating:

> "Which statements are labelled 'true' or 'false' in a given community depends on its prevailing exemplars, interests, and goals. Changes in any of the latter lead to changes in the former. The sorting of beliefs or statements into 'true' and false statements' is thus done relatively to prevailing exemplars, interests, and goals" (Kusch 2002: 270).

Acknowledging "[...] we cannot escape our own contingency as members of our own culture and its tradition" (Kusch 2002: 270), however, does not imply that all statements are only relatively true. Differentiating these two notions of the word "relative" as "nearly" versus "in relation to" may be crucial for sorting debates around relativism as Kusch has argued convincingly. Kusch's embrace of relativism goes too far for many, if not for all social epistemologists portrayed before. But - for better or worse - I consider his arguments for such a relativist stance to be consistent.

6.3.5 Methodological Issues: Naturalism and Quantification

Methodological decisions are another criterion in which social epistemologies differ. Although it seems that debates about methodology are less vivid than other discussions, I would argue that these methodological choices are fundamental and have considerable effects. Two issues here are of crucial concern: naturalism in epistemology and the role of quantification and stochastical methods. At first I briefly address the topic of naturalism more broadly conceived as the receptiveness to empirical data before turning to the issues of quantification and the use of stochastical methods in particular.

Naturalism in Epistemology

Several social epistemologists portrayed consider their approaches to be naturalistic in the sense of being receptive to *empirical studies* of knowledge and epistemic practices. Miriam Solomon even considers naturalism to be the "most important development in epistemology over the last fifty years," and claims that her "arguments for social empiricism have been consistently naturalistic" in that [she] appealed to case studies and to the empirical sciences of knowledge (social psychology, cognitive psychology, sociology, etc.) in making them (Solomon 2001: 137). The basic idea is that in order to propose how epistemic practices *ought* to be conducted, one has to know how they *are* actually conducted. Hence, any attempts to be normative or prescriptive about epistemic practices should be rooted in the analyses of actual epistemic practices. Yet while the role of psychology and cognitive science for epistemology has long been acknowledged in contemporary analytic philosophy, the openness to other empirical disciplines, such as sociology, anthropology or history, has been more circumspect.

In principle, those authors who consider their approaches to be naturalist have imported either insights or methods from the empirical sciences into philosophy. Various authors have used historical case studies from the history of science for the development of their theories. For instance, Paul Thagard, whose considerations on collaboration in science I portray in Chapter 9, has not only conducted extensive case studies in the history of science and medicine (Thagard 1998a; Thagard 1998b), he has also used computational methods and modeling to amend philosophical methods (Thagard 1998c). Longino has developed her own approach based on analyses of theories of human evolution and of the hormonal basis of sex differentiated behavior (Longino 1990). She has also thoroughly analyzed the challenges that sociologists, anthropologists and historians of science have posed to philosophy of science (Longino 2002c). Longino specifies that her approach is naturalistic in the sense that "[...] it treats the conditions of knowledge production by human cognitive agents, empirical rather than transcendental subjects, as the starting point for any philosophical theory of knowledge, scientific or otherwise. It is not naturalist in the sense of treating knowledge as a natural kind whose nature or essence can be discovered (by scientific or philosophical methods)" (Longino 2002c:

10). I share Longino's position concerning the necessity for social epistemology to be naturalist in focusing on actual epistemic practices and to take insights from different empirical accounts of knowledge and science seriously while rejecting the idea that knowledge is a *natural kind*.

As should have become obvious, I do share a commitment to naturalism when understood as the receptiveness to insights of knowledge and epistemic practices from the empirical sciences. I do also consider quantitative methods, various stochastical techniques in particular, to be important for social epistemology. Indeed, in the second part of this thesis I outline a type of epistemic sociality that is entirely based on statistical aggregation of data and defend it as an important socio-epistemic mechanism. However, I am also aware of the inherent dangers of quantifying the qualitative. As Friedman and Nissenbaum have noted in their analyses of information systems, the formalization of human constructs is always bound to biases, quantifying the qualitative always means losing some information while focusing on other information (Friedman and Nissenbaum 1997). This insight will remain important in my analyses that follow in the succeeding chapters.

But what is even more dangerous is that quantification may put a patina of scientific integrity atop faulty assumptions. Proclaiming the use of multivariate statistical methods or recommending the application of Bayes' Theorem for judgment aggregation may give social epistemology a more scientific ring—numbers are often convincing—yet one of the crucial mantras here should be: garbage-in, garbage-out. Quantitative results can only be valid if the data that feed into them are valid. Hence, paying close attention to the *methods of quantifying the qualitative* and *the quality of the initial input data* is paramount for the assessment of quantitative approaches in social epistemology. If the basic assumptions, the variables or the input data are faulty, the results are at best useless but, very often, misleading.

Below, I critically assess two attempts at quantification that have been proposed within social epistemology. Initially, I assess Miriam Solomon's *Social Empiricism* in some detail, with respect to her stance towards naturalism in general and her mode of quantification (the "improper linear model") in particular. I argue that Solomon uses quantification inappropriately and that such a usage is best avoided. Following on this, I look at Alvin Goldman's Bayesian account of testimony and argue that quantification is dangerous here as well because it is likely to conceal the qualitative and highly subjective nature of the input variables.

Quantification and the Use of Stochastical Techniques in Social Epistemology

Naturalism and Quantification in Solomon's Social Empiricism

In addition to the scholars I have already mentioned, Miriam Solomon has also analyzed episodes from the history of science to develop her theory of *Social Empiricism.* She claims that her *decision vectors* are not conceptual but empirical, because they are derived from these case studies. Moreover, she frequently alludes to the relevance of statistical methods or multivariate models for the development of appropriate frameworks for understanding scientific change. Although I welcome Solomon's reference to historical case studies and stochastical procedures, I am not convinced by her realization. Below I address some of the methodological issues concerning the origin of Solomon's decision vectors, her classification into empirical and non-empirical decision vectors, and her usage of an "improper linear model".

Origin of Decision Vectors

Solomon lists a number of decision vectors in her theory including: ideology, pride, conservativeness, radicalism, elegance, simplicity, representativeness heuristic, competitiveness, peer pressure, defense to authority, birth order of scientists, etc. These are all examples of non-empirical decision vectors. Empirical decision vectors are, for instance, salience and the availability of data, or an egocentric bias towards one's own data or a preference for a theory which generates novel predictions (Solomon 2001: 57f). Even at first glance these decision vectors look quite diversified. But as Solomon reminds us - "a decision vector is *anything* that influences the outcome of a decision" (Solomon 2001: 62). One would assume that given this broad definition many things would qualify as decision vectors in science. Indeed, Solomon criticizes other social epistemologists for sidestepping complexity by considering too few decision vectors (e.g. Goldman 1992, Kitcher 1993). To her mind, "the number of types of decision vectors is probably between 50 and 100" (Solomon 2001: 62).

But how does Solomon arrive at this number? And how does she arrive at the distinct decision vectors she labels "ideology" or "pride" or "birth order"? The decision vectors that Solomon lists appear to be a quite random synopsis of various social, political, motivational and cognitive biases extracted from a variety of theoretical and empirical literature (Solomon 2001: 51ff). How exactly she ends up with the proclaimed 50-100 discrete vectors is not obvious. Throughout the book, there is no method discernible despite a seemingly random choice of labels for different factors that might or might have not have had an impact on the development of science. While the listing of decision vectors from various sources is a valid procedure of course, problems occur when these decision vectors get *quantified*. I outline this problem below when addressing Solomon's usage of an improper linear model.

Empirical versus Non-empirical Decision Vectors

Before, doing so, however, I want to return to Solomon's differentiation between empirical and non-empirical decision vectors. Since this is the primary classification of decision vectors and forms the conceptual basis of her Social Empiricism, the plausibility and discriminatory power of this differentiation is crucial. Surely, while some factors which influence decision making in science are related to empirical success, I am not convinced by Solomon's examples. As compared to the extensive list of non-empirical decision vectors, Solomon lists only four empirical decision vectors: salience of data, availability of data, egocentric bias towards one's own data, and preference for a theory which generates novel predictions (Solomon 2001: 57).

But are these really empirical decision vectors, i.e., are they necessarily linked to empirical success? For instance, are salience and availability of data necessarily empirical decision vectors? Or could they also be non-empirical? Just imagine that some very important literature on a certain topic exists only at a library to which you either have no access or knowledge of. Here data would not be available or salient, even though it exists, yet could be highly relevant for future empirical success. Alternately, think about the decisions that are made within an epistemic community – or maybe even within a research group, about what literature and which methods should be used. In this case, information may simply not be salient to someone because it is not considered to be important *for reasons other than empirical adequacy*. Or – to link this to more recent developments: what happens to data, information, articles, etc. that are not available in electronic format? Think about how many pages of Google results one can check when searching for information, and how much relevant information is neglected for reasons other than empirical adequacy? Research on information retrieval suggests that these sources of information, however valid they might be, are neglected precisely because they are not considered salient when not easily accessible. To put it in a nutshell: saliency often depends on other non-empirical decision vectors. Thus, neither are decision vectors independent, nor is the differentiation between empirical and non-empirical decision vectors indisputable.

The Improper Linear Model: Simple Addition of Independent, Equipollent Factors?

Solomon proposes an 'improper linear model' as the simplest form of multivariate analysis for the basis of her Social Empiricism. The terms "improper linear model" and "multivariate analysis" sound quite mathematical, so what exactly is an improper linear model? Solomon explains that "[i]n this analysis, variables are not assigned their actual magnitudes nor are their interactions with one another considered. It is simply noted whether the effects are positive (+) or negative (-), and then the effects are summed additively" (Solomon 2001: 77). Note, however, that this model states nothing about the origination of its variables. It is only an assertion that once variables have been defined, the improper linear model will allow for the simple addition of such variables. The fundamental flaw of Solomon's approach to my mind lies in *applying this quantitative model to the qualitative decision vectors extracted from various sources*.

Solomon has extracted her list of decision vectors from a wide variety of literature on social, political motivational and cognitive biases. As noted before, these include a variety of very different factors from large-scale political factors like ideology to theoretical values, such as simplicity or elegance, to peer pressure and competition in science. While I agree that all these factors can play a role in decision making processes in science, it is the next step in Solomon's approach that is fatal: she counts her decision vectors.

For instance, when having analyzed the distribution of decision vectors for the Continental Drift Dispute between 1920 and 1950, Solomon summarizes the relationship of empirical and non-empirical decision vectors for the three theories as follows:

Empirical Decision Vectors	
Permanentism	+2
Contractionism	+2
Drift	+1
Non-Empirical Decision Vectors:	
Permanentism	+5
Contractionism	+4
Drift	+4

Table 3: Distribution of Decision Vectors (Solomon 2001: 91)

If one adds up all the possible factors that have had an effect on the distribution of epistemic labor, and on consensus and dissent within the scientific communities working on plate tectonics from 1920 till 1950, one arrives at following conclusion: two empirical decision vectors worked in favor of Permanentism and Contractionism, and one worked in favor of Drift. Five non-empirical decision vectors worked in favor of Permanentism, four in favor of Contractionism and Drift. Solomon concludes that "this is a fairly equitable distribution of empirical decision vectors, and equal distribution of non-empirical decision vectors, indicating appropriate distribution of research effort" (Solomon 2001: 91).This distribution is the reason why there was normatively appropriate dissent in the Continental Drift Dispute between 1920 and 1950. The distribution of decision vectors then shifted between 1958 and 1970, and consensus was formed on drift. Solomon asserts that this consensus occurred because the distribution of decision vectors shifted in a certain way and that consensus on plate tectonics was normatively appropriate in this later time period. Here I do not want to address the question of whether this consensus really was normatively appropriate (for a critical view, confer Oreskes 2008). Rather, by taking a closer look at what exactly has been calculated here I want to emphasize the inadequacy of adding up decision vectors.

Here are some examples of factors that were counted as "1 decision vector" in Solomon's innocuous table (cf. Solomon 2001: 87ff).

- European geologists' tradition of theorizing and speculation: 1 non-empirical decision vector in favor of Drift and Contractionism.
- Anti-German Feelings: 1 non-empirical decision vector against Drift
- Birth Order of Scientists: 1 non-empirical decision vector in favor of Drift, Contractionism and Permanentism

- Belief Perseverance Phenomena: empirical decision vector: two points for Permanentism, one for Contractionism.

I do not even want to dwell on the question of why belief perseverance is an empirical decision vector and how it is that it can count double if alleged variables are not assigned magnitude in an improper linear model. I find the way in which random qualitative factors are arithmetically calculated highly irritating. Just to give another example: in the case of cancer virus research, Solomon counts the fact there was a lack of fear of contagion as *one* non-empirical decision vector and the fact that the main advocate of the research was an immigrant as *another* (one!) non-empirical decision vector—each of which have to be set against one other.

While I do certainly agree that all these factors play a role, quantifying them as *discrete decision vectors* to be measured against one another is an astonishing simplification. How could something as broad, complex and pervasive as ‘conservatism’ or ‘belief perseverance’ be condensed into one single decision vector? Moreover, how can one reasonably assume that all the decision vectors have the same weight?

Solomon’s simple addition of decision vectors is based on two premises outlined in the definition of the improper linear model above. First, decision vectors are not related. Second, they do not differ in weight - if there is a difference in magnitude, it is neglected. Both assumptions are highly problematic, but only if they are fulfilled can an impropriate linear model be applied and decision vectors added up.

Is it reasonable to assume that the decision vectors in Solomon’s model are *unrelated to each other* and of the *same strength?* I start with the question of whether all vectors have the same strength. Please note that a simple, unweighted addition of decision vectors *implies* that strength does not matter. This means that all of the above mentioned factors – ideological issues, pride of scientists, simplicity of the theories, competitiveness and peer pressure within scientific fields, birth order of scientists, etc. – have the valence of 1. Hence all factors have the exact same impact on scientific decisions, on consensus and dissent within a scientific field. Does this sound plausible?

Furthermore, the longer one looks and the more one reads about certain episodes in the history of science, the more decision vectors one may be able to extract. Remember that according to Solomon, decision vectors are *anything* that affects the outcome of decision making in science. *Anything*! How can one possibly be sure that she has not missed something that has affected the outcome of a scientific debate? If strength of the impact is irrelevant – as must be the case if Solomon's improper linear model is to be applied – then this process is, in principle, endless and *each distribution is an arbitrary termination of counting*. The assumption that the strength of decision vectors is irrelevant seems not only implausible, but also leads to unwelcome consequences concerning the arbitrariness of the distribution of decision vectors. This alone should be reason enough to abandon such an improper linear model when analyzing the role of different biases in the history of science.

The second prerequisite, that decision vectors must be independent, is also highly implausible in the case of such biases. Isn't it rather unlikely that there is a relationship between pride, agreement to scripture, competitiveness and peer pressure? This problem of interdependence becomes even worse if one looks at the concrete historical case studies which Solomon analyzes. Not only are numerous decision vectors related, but her decision whether to label something as a distinct decision vector, and is therefore calculable, is subjective and arbitrary. This leads back to the question of how Solomon arrives at her decision vectors in the first place. Her list of decision vectors appears not only to be an arbitrary synopsis of various heterogeneous influencing factors, but the application of these decision vectors to historical case studies is also highly subjective. To be sure, I consider neither the heterogeneity of the factors nor the subjective attribution in itself problematic. But the moment one starts counting these heterogeneous, subjective variables, one commits a fatal mistake: the data fail to satisfy the requirements for an improper linear model as I have shown. One also obfuscates the subjectivity and arbitrariness of the input data and puts a patina of scientific integrity over questionable or even faulty assumptions. Clearly, such as method does not seem recommendable to make claims about the appropriateness of scientific conduct or to even guide science policy.

Alvin Goldman's Bayesian Account of Testimony

Alvin Goldman's account of testimony is another example of quantifying the qualitative within social epistemology. As was outlined in the previous chapter, Goldman recommends the application of Bayes' Theorem to testimonial evidence as a socio-epistemic practice that raises the likelihood of obtaining true belief. Remember that the central task for Goldman's veritistic evaluation of socio-epistemic practices is to identify those practices that lead to largest increase in truth known in a community. With respect to testimony, Goldman seeks to identify veristically superior testimonial practices of testifier and recipient. The goal would be a general strategy for accepting or rejecting testimony which is valid in all situations. Yet since in different reporting environments, different acceptance strategies concerning testimony may lead to the best results, many simple heuristics do not work. Hence, Goldman proposes Bayesian inference as such a general strategy.

As noted before, applying Bayesian inference to testimonial evidence only works, if the following probabilities are known: the prior probabilities of an event x and not-x, as well as the conditional probability of the witness testifying x given that x has occurred and given that x did not occur.[106] These latter conditional probabilities imply an assessment of the competency and honesty of a witness. Moreover, all of these values have to quantitative, for Bayes' Theorem to be applicable. My criticism of Goldman's proposal concerns exactly this quantitative assessment of the competency and honesty of the testifier which is a necessary input variable, this *estimation of testimonial likelihoods.* The crucial question is how we are supposed to arrive at quantitative values to indicate the perceived competency and honesty of our fellows and friends?

Given the centrality of the correctness of such testimonial likelihoods, it is surprising how little Goldman has to say about how to arrive at correct likelihood ratios and it is precisely this central blind spot that has lead to a US Supreme Court verdict arguing *against* the use of Bayesian inference in court. Kusch reports on this verdict of the Supreme Court arguing that the likelihoods of different pieces of evidence are a result of judgment and that the quantified results and mathematical processing of this initially qualitative data "[...] might easily 'conceal the element of judgement'" (Kusch 2002: 80f).

Clearly, the merits of Bayes' Theorem are not to be questioned in this thesis. Indeed, Bayes' Theorem is not only of great value for the empirical sciences, it has also been successfully employed in many social software applications. What I do question is the utility of recommending Bayesian inference for processing testimonial evidence. First of all, only few people know how to calculate it and it also quite unlikely that even if they

[106] For further information on Bayes' Theorem please confer the section on Goldman in the previous chapter.

do, they will employ this method on all those innumerous occasions in which they accept testimonial evidence in their daily live. One may argue that Web applications which make use of Bayesian inference are exactly an example of a useful distribution of cognitive labor between humans and computers. I return to these issues in the second part of this thesis when assessing the merits and dangers of algorithmic authority. However, the second problem is even more fundamental. Using Bayesian inference is only possible if the testimonial likelihoods are known. Even if we can delegate the processing to machines, we need to feed in the data about the competency and honesty of the testifiers – and this is a highly subjective judgment. If testimonial likelihoods are not know and can only be approximated, guessed or even chosen randomly, then making use of Bayesian inference is likely to obfuscate the amount of subjectivity and arbitrariness of input data.

6.3.6 Normativity

Almost all social epistemologists stress that their theories are normative. Normativity here comes in two forms. On the one hand it means that a social epistemology is able to distinguish knowledge from mere belief, to distinguish what *is true* from what is *taken to be true*. On the other hand, normativity refers to *prescription*, to recommendations on how to *improve* socio-epistemic practices.

Normativity1: Criteria for Knowledge

The ability of an epistemology to distinguish knowledge from mere belief seems a crucial requirement for Anglo-American social epistemology and is closely related to some of the debates outlined before: the acceptance or rejection of the definition of knowledge as justified true belief, the reception or rejection of sociological theories, the stance towards relativism, etc. Being able to distinguish *knowledge as true belief* from *mere belief* is of particular concern for Alvin Goldman's veristic epistemology. It is also the rationale behind characterizing Steve Fuller's and Steven Shapin's work to be "social doxology" (Goldman 2003: 7), the reason for denying that "revisionist social epistemologies", rooted in postmodernism, deconstructionism, social constructionism, and various social studies of science, including the "strong programme" in sociology of science " are *real* social epistemologies (Goldman 2009b).

That the distinction between knowledge and mere belief is crucial for Goldman's veristic account should be of little surprise. Yet even a contextualist epistemologist like Helen Longino asserts that although knowledge on her account is always only partial, plural and provisional, her contextual critical empiricism can nonetheless distinguish knowledge from mere belief (Longino 2002c). To support this claim Longino analyzes two examples of epistemic practices which any normative social epistemology should discredit by showing that they are not epistemically acceptable: fortune telling and defending creationism. Remember that Longino defines a cognitive community as a group bound by a "set of common goals and shared public standards regulating critical (knowledge-productive) discourse and the stabilization of representations as knowledge" (Longino 2002c: 145). Crucially, those communities determine the standards by which the evaluate knowledge claims and depending on the community, the context or the epistemic goal, these standards may differ. If all communities can make up their own standards, how can it be argued that creationists and fortune tellers are not creating knowledge? According to their own standards their epistemic practices are epistemically acceptable – so according to which standards do they fall short?

Lognino argues that there are two crucial requirements which cognitive communities have to fulfill to be normatively appropriate. First, they must be willing to make their standards subject to scrutiny. Second, cognitive communities must have an interest in the accurate description of their physical environment (Longino 2002c: 162). She demonstrates that in the cases of creationism and fortune tellers these two prerequisites are not met. While creationists do not open their central dogmas to criticism, fortune tellers are not interested in depicting an accurate description of reality when reading tea leaves. Both aspects however, are necessary conditions for epistemic communities and the pursuit and attribution of knowledge.

Martin Kusch goes furthest in arguing that normativity can only be understood in relation to epistemic communities and not as an interaction between the world and the individual mind (Kusch 2002: 121). One of the central premises of Kusch's *Communitarian Epistemology* is that "normative phenomena—rules, norms, conventions, prescriptions, and standards of correctness—can exist only within communities (Kusch 2002: 175). Drawing on arguments against *private languages,* Kusch argues for what he calls the "Strongest Present-Tense Community Thesis" according to which "[a]n individual is able to follow a rule only if the individual is currently a participating member of a group in which the very same rule is followed by other members (Kusch 2002: 181).

Normativity2: Improving Socio-Epistemic Practices and Systems

The second denotation of normativity refers to prescription, to recommendations on how to improve socio-epistemic practices and systems. Examples of this form of normativity are Steve Fuller's (Fuller 1988) and Miriam Solomon's (Solomon 2001) attempts to influence science on a macroscopic level, Helen Longino's (Longino 2002c) social norms for social knowledge and Goldman's recommendation of Bayesian inference to assess testimonial evidence (Goldman 2003). Although these recommendations target different entities and operate at different levels, the goal of all the recommendations is the same: to change the parameters of socio-epistemic systems and socio-epistemic practices for better. I do not rehearse all normative-prescriptive epistemologies at this point since they have been portrayed in some detail before. Their implications, possible implementations and utilizations in socio-technical epistemic systems are addressed below in the section on social epistemology and social software.

In contrast to all other comprehensive social epistemologies portrayed in Chapter 5, Martin Kusch's *Communitarian Epistemology* does not provide any normative-prescriptive framework. Indeed, Kusch argues that his interest lies in understanding rather than changing socio-epistemic practices (Kusch 2002: 2). Kusch is highly skeptical about social epistemologists who confine themselves exclusively to normative issues and deplores that while "social epistemologists eagerly set norms for groups and interacting individuals [...] they never stop to ponder the nature of norms, standards, institutions, or groups" (Kusch 2002: 115). In other words, he criticizes the insufficient reception of empirical analyses from the social science and argues that this neglect is particularly astounding given epistemologists' interest in cognitive psychology or artificial intelligence. This topic leads back to the debates around naturalism outlined before: while many social epistemologists claims to be naturalist, their interpretations differs profoundly as to what this implies in general and which empirical fields of research are considered to be relevant in particular.

While I agree with Kusch on the necessity to analyze actual epistemic practices before making recommendations on how to improve them, I consider this normative-prescriptive orientation to be a crucial aspect of any social epistemology. Given the breadth and depth of his analyses and his own proposal for a communitarian epistemology it would have been interesting to see which normative-prescriptive conclusions Kusch draws from them. Especially given his indebtedness to feminist epistemologies I find this neglect somewhat disappointing. In the end he assures however, that there is a place for a normative-prescriptive epistemology within *Communitarian Epistemology* and that he does not question its usefulness (Kusch 2002: 285). Yet against the background of his *Communitarian Epistemology* it should not come as a surprise that "[...] communitarian epistemologist insists that all such normative—prescriptive endeavors have their roots in local contexts and contingencies" (Kusch 2002: 285).

Normativity1 & 2: Criteria for the Critical Assessment and Amendment of Socio-Epistemic Practices and Systems

To conclude, I argue that social epistemologies should not remain purely descriptive but that a normative stance is crucial. If one wants to critically assess socio-epistemic systems and practices, as I do in the second part of this thesis, such a normative orientation is indispensable. Without a normative stance, there are no criteria for critical assessment. Yet I share Longino and Kusch's assertion that normativity and hence criteria to assess epistemic practices and systems vary between different communities as well as within communities depending on the context and the goals of epistemic inquiry. Different criteria are of different relevance depending on the current norms and the goals of epistemic inquiry. Alvin Goldman has proposed five general standards of epistemic appraisal (Goldman 1992: 195): reliability, power, fecundity, speed and efficiency. Although these criteria seem to be a good starting point for the analysis of socio-epistemic practices as well as socio-technical epistemic systems, they are by no means the only criteria imaginable. Other criteria may be heuristic fruitfulness, freedom from bias or accountability (e.g. Friedman and Nissenbaum 1997, Nissenbaum 1997). As long as these criteria have epistemic relevance, they are valid criteria for a socio-epistemological analysis of epistemic practices and systems.

Critical assessment however does not have to be the final step for social epistemologists. Rather they should develop and propose alternatives for those socio-epistemic practices and systems that they consider inadequate or wanting. The means and possibilities to *improve* socio-epistemic practices are as manifold as the criteria for assessing them.

With respect to socio-technical epistemic systems a wide variety of interventions and recommendations can be imagined. There could be guidelines on how to avoidance different forms of bias or on how to provide access to information for more people. Another possibility could be to support processes of knowledge creation by offering platforms for critical discourse, by helping to find partners for epistemic endeavors. Social epistemology with normative aspirations has to come up with ideas of how to support the creation as well as the distribution of knowledge and there are innumerous aspects which may be taken into account. Exploring them should be a central task for social epistemologists – a task that is taken up in the third part of this thesis.

6.4 Social Epistemology Beyond Science: Social Epistemology for Social Software

All social epistemologists portrayed use science as a field of reference for developing their social epistemologies. However, while some restrict their social epistemologies to apply to science only (e.g. Solomon 2001), others have proposed social epistemologies that either describe more general models of the sociality of knowledge (e.g. Kusch 2002) or have explicitly addressed socio-epistemic practices beyond science (e.g. Goldman 2003, Goldman 2008, Thagard 1997b). Since the goal of this thesis is to provide a model to analyze socio-epistemic practices and systems on the Web those social epistemologies that go beyond science or even directly address information and communication technologies may appear to provide more appropriate frameworks. However, I argue that even from those approaches that are as decidedly "science-only"-models, such as Solomon's *Social Empiricism* or Longino's *Critical Contextual Empiricism*, important insights for the analysis of socio-epistemic practices and systems on the Web can be extracted. In the following I briefly outline the relevance of different social epistemologies for the analysis of epistemic social software. Afterwards I portray some of the initial attempts to analyze information and communication technologies (ICT) from a socio-epistemological perspective. In particular, I refer to the Alvin Goldman's, Paul Thagard's and Don Fallis' analyses of ICT as well as a recent *Episteme* special issue on epistemic mass collaboration. I conclude this chapter with a brief summary of my own position on the issues raised in this chapter in so far as they are relevant for the developments of my own socio-epistemological model to be depicted in the next chapters.

6.4.1 General Models of Epistemic Sociality

Martin Kusch proposes his *Communitarian Epistemology* as a general social epistemology that is not restricted to scientific knowledge creation, because according to him "[…] as far as their basic 'socialness' is concerned, scientific and ordinary forms of knowledge do not differ from one another" (Kusch 2002: 4). I agree that depending on the level of abstraction socio-epistemic practices on the Web share fundamental aspects with socio-epistemic practices in science.This openness of his approach towards non-scientific knowledge and more secular epistemic practices is clearly crucial for the questions of this thesis. If we aim at understanding knowledge and epistemic practices in epistemic social software or socio-technical epistemic systems in general, our analyses will indeed be hampered if we focus too narrowly on scientific, and especially only experimental forms of knowledge creation. Hence, my own socio-epistemological framework also aims at providing a general model of epistemic sociality. It is based on analyses of socio-epistemic practices on the Web and on my reception of social epistemologies which are mostly based on the analyses of socio-epistemic practices in the science, but due to the abstractness of the types of epistemic sociality I propose, this framework can be considered a general socio-epistemological model which is applicable to a wide variety of socio-epistemic practices and systems.

6.4.2 Social Epistemologies for Science and Their Implications Beyond Science

Miriam Solomon's as well as Helen Longino's approaches have been either developed in the context of science (Longino 2002c) or decidedly for science (Solomon 2001). However, I would argue that even from those approaches important insights for the analysis of socio-epistemic practices and systems on the Web can be extracted. Despite my critique of Solomon's approach, I nonetheless consider her focus on the community as the decisive level at which scientific rationality gets decided interesting for socio-epistemological analyses. Such a macroscopic perspective surely can provide important insights into the socio-epistemic processes on the Web. Also, her list of possible decision vectors may provide a starting point for critical analyses of socio-epistemic practices and systems and the role of biases therein. The aspect that I consider most important in Solomon's as approach concerns her macroscopic view on the scientific field as a whole as well as her emphasis on the benefits of epistemic diversity. Various forms of distributing labor over multiple agents on the Web as well as the acknowledgement of the diversity of their motives are crucial for any general framework for assessing socio-epistemic practices and systems.

Helen Longino's *Critical Contextual Empiricism* has also been developed within the framework of scientific knowledge creation. Nonetheless, I consider her approach to be exceptionally inspirational and fruitful for the analyses of socio-epistemic practices and systems more broadly conceived. One aspect that I find particularly useful in Longino's account, are her four social norms for social knowledge. To my mind, those social norms can and should be applied to assess and amend socio-epistemic practices not only in science but also in other epistemic systems. Indeed, I would argue that certain developments in information and communication technologies can even ease the implementation of these social norms. Take for instance the first requirement for her transformative criticism: the demand for publicly acknowledged venues for criticism. This demand implies actually two needs: criticism need more space – in journals, conferences, etc. and it has to be valued to a greater extent, maybe even to the same extent as original research. In science, original contributions are usually considered to be far more important than critiques of the works of others and are therefore given more space in journals. While reference to others clearly plays a role in publications and therefore criticism is implicitly implied in scientific conduct, the space that is given to explicit criticism is clearly smaller and less renowned than the space given to original research. Also, pure replications of experimental results are rarer than one would assume given their alleged relevance for science. Original works are considered to be more important and the reputation of a scientist depends far more on his original research than on the reviews her writes about the works of other people. Just think about the perceived value difference between a journal publication providing new content and a book review.

However, if we take not only Longino's approach serious, but also the more general idea that science in principle consists in mutual criticism and improvement, than it should be obvious that constructive and fruitful criticism is just as central to the development of science as the production of new data. This emphasis on the crucial role of effective criticism becomes important in Part 3 of this thesis again. Especially in an environment in which space is scarce, the perceived value difference between original contributions and replications or critical assessment has negative consequences. If you only have a certain number of pages for a given journal edition, most editors would probably try to include as much new content as possible and this takes away space for criticism. Given that such space constraints cease to exist on the Web, the rehabilitation and encouragement of criticism is a crucial epistemic task that should be of vital concern when developing new socio-technical tools for science and beyond.

Or take the fourth prerequisite of tempered equality. The Web and social software in particular have provided an opportunity to include more voices into epistemic communities. Think of Wikipedia. Think of the open source movement. Of course one can argue that the chances for people to take part in epistemic communities are not equal for everyone due to the digital divide and that there is a gender imbalance in the open source community. Nonetheless, joining epistemic communities, taking part in various epistemic endeavors, getting into intellectual exchange with others has clearly become easier with the advent of the Web and other information and communication technologies. The question of *how* to temper equality also had to be – and has been tackled in different applications and a multitude of different possibilities has been tested.

The Wikipedia predecessor Nupedia for instance relied on academic credentials to select contributors – and failed with this strategy. Its successor Wikipedia then opened the door and allowed for the participation of everyone who had access to the Internet. Only once the community grew they introduced a hierarchy of users by differentiating normal users from administrators with more rights. Slashdot.org would be an example of a website that works quite different in the selection of content. All users are permitted to post comments on this platform and other readers can rate these comments. These ratings are aggregated into collective judgments about the quality of these comments and those can then used by the other users to filter comments (cf. for instance Benkler 2002). As such slashdot.org and other similar applications combine human judgment with various algorithms to evaluate content, and by doing this "temper" the intellectual authority of the content providers.

Given that Longino argues for local and situated ways of tempering intellectual authority and a consideration of the actual epistemic practices, a fruitful collaboration between epistemology and web science seems feasible. Epistemology can be of use for the design of such systems by epistemically assessing the modes of credit attribution in such systems. They can also help in developing new approaches. A closer analysis of the mechanisms involved however, should also not leave epistemology untouched. The different evaluative mechanisms embedded in these socio-technical systems are not only of interest for the empirical assessment of epistemic practices that should form the basis of normative-prescriptive social epistemologies. They can clearly also deliver inspiration for new epistemic accounts of how to *rationally attribute intellectual authority*. I would therefore conclude that Longino's approach makes a clear case for the potential fertility of applying social epistemology to research on and development of social software. I come back to the issues raised here in much more detail in the third part of this thesis.

6.4.3 Socio-Epistemological Analyses of Information and Communication Technologies

So far, there have been only few attempts of social epistemologists to address information and communication technologies. Alvin Goldman (Goldman 2003) and Paul Thagard (Thagard 1997b) have recognized the relevance of those technologies for knowledge and science in particular already in the 1990ies. Although there have been dramatic changes in ICT since these days, some general aspects of Goldman's and Thagard's analyses clearly remain valid. Moreover, Goldman has continued working on the topic and proposed a social epistemology of blogging more recently (Goldman 2008). In 2006, Don Fallis has provided an extensive review on the relationship between social epistemology and information science (Fallis 2006) and in 2008 he edited a special issue of the journal Episteme on "The epistemology of mass collaboration", in which numerous aspects of Wikipedia, the relevance of decision markets and the differences between Web2.0 and the Semantic Web were outlined. I portray these approaches in same detail below since they represent some initial attempts to assess information and communication technologies from a socio-epistemological perspective. I conclude this portrayal by arguing that while these approaches are interesting they can only be considered starting points for any comprehensive understanding of the socio-epistemological relevance of such systems. Moreover, the majority of these approaches are limited by a too narrow focus on the distribution of knowledge while neglecting their relevance for knowledge creation.

Thagard’s Internet Epistemology

Thagard’s (Thagard 1997b) *internet epistemology* is an early example of how to assess the epistemic merits and problems that come with the usage of Web for epistemic purposes. Clearly, not only the Web applications themselves, but also *research on ICT* has developed profoundly since 1997. A whole range of organizations and journals have emerged that are devoted to an assessment of the Internet and the Web since then.[107] Nonetheless, I do think that Thagard’s analyses are still interesting. Especially his portrayal of “a day in the life of a cyberscientist” is pretty close to the day of many scientists today and did not remain just “speculative science fiction”(Thagard 1997b).

Thagard portrays a short history of the Internet and the Web and depicts the role ICT plays in some major science projects. Amongst the main examples he chooses are two large-scale scientific collaborations: CERN and the Human Genome Project, the former being the birthplace of the Web. Comparing the Web to the advent of the printing press, he makes use of five general epistemic standards proposed by Alvin Goldman (Goldman 1992: 195) to assess the merits of ICT. These five standards of epistemic appraisal are reliability, power, fecundity, speed and efficiency. He concludes his analyses on the contributions of ICT to science in a table and instead of describing them in length, the table is depicted below.

[107] One important association is the Association of Internet Researchers (AoIR). And a query SULAIR, Stanford University’s database for electronic journals and newspapers resulted in 218 journals that have the term Internet in their title, not to mention those entailing “new media”, “Web”, etc.

	Reliability	Power	Speed	Fecundity	Efficiency
Email, News Groups	feedback for corrections	many answers available	faster than mail	multiple recipients	cheaper than paper mail
Hypertext	easily revised	follow links, use search engines	instant publishing, no wait for access,	searching widely available, distance irrelevant	storage cheap
Animation, Video, VRML	more accurate depiction of structures and motion	lots of visual information not otherwise available			
Java	software not under local control	instant provision of software to do examination, searches	no wait for software	use by everyone regardless of kind of computer	no need to buy software, or spend time on getting it
Databases	updatable, checkable	huge amount of information available	fast searchers, instant availability	accessible to many	storage is cheap
Preprint Archives	potentially quick feedback	find out latest research results	instant access	journal access unnecessary	total cost much lower than print
Conferencing	immediate corrections	combine new ideas	no need to meet	everyone involved	cheaper than meeting

Table 4: Internet Technologies and Scientific Research (Thagard 1997b)

Thagard concludes that while the Web and the Internet clearly led to increases in power, speed, fecundity and efficiency, the effect on reliability is less clear. Reliability understood as the "ratio of truths to the total number of beliefs" can be increased as well as decreased on the Web. While he asserts that in the end, the question of whether reliability will be increased or decreased will depend on the "users' intellectual tools for discriminating between reliable and reliable sources of information", taking a look at the first column of table above is also instructive. Almost all the benefits of the Webwith respect to reliability are related to the Web's ability to *increase mutual criticism*. Feedback, revisions, local control, updatability, and verifiability are the major factors for reliability. It is only a short mental leap to relate them to Helen Longino's four standards to enable transformative criticism (Longino 2002c) and only another small step to assess Web2.0 projects such as Wikipedia with respect to their ability to foster and support such criticism. Indeed, Thagard stresses this relationship between science, criticism and ICT himself by stating that "[s]cience, like knowledge in general, is an inherently social enterprise in which achieving truth and avoiding error gains enormously from feedback that Internet technologies can help to provide" (Thagard 1997b).

I consider Thagard's usage of Goldman's five epistemic standards to assess the merits and problems of ICT to be instructive and think that these standards can be also used for the assessment of epistemic social software, although they need to be amended. From those five standards, the question of reliability might be the most interesting and tricky.

Alvin Goldman's Veristic Analyses of ICT

Alvin Goldman applies his social epistemology to various fields beyond science as was outlined in the previous chapter. Not only does he analyze socio-epistemic practices in education, law and democracy, he also devotes a whole chapter of his book "Knowledge in a Social World" to "The Technologies and Economics of Communication" (Goldman 2003).[108] Many of the issues which Goldman identifies at the end of the last century, are still valid today although the Web has changed profoundly since then. Indeed, some of the topics around information retrieval, the utility of autonomous agents or the relevance of ICT for scholarly communication are probably even more hotly debated today than at the time when Goldman published his book. Moreover in a more recent paper, he analyzes the social epistemology of blogging. More specifically, he focuses on the role of blogging for democracy. The major question for him is whether the Web "[...] is better or worse in epistemic terms than the conventional media, in terms of public political knowledge generated by the respective communication structures"(Goldman 2008: 12).

His arguments in this recent paper are consistent with the veristic perspective expressed in "Knowledge in a Social World" (Goldman 2003). For instance, he argues that conventional media often have the positive function of filtering information and compares this process to peer review in the sciences. With respect to an overall veristic analysis of blogging, Goldman argues that in the end, the veristic impact of the blogosphere will depend on the users' motivations. Some users may use the abundance of unfiltered information to pick out only the information that corroborates their views, which are possibly biased or even utterly wrong.[109] Others however, might use the

[108] Please confer the section on Goldman in the previous chapter.
[109] Such an information retrieval behaviour is outlined for instance in Cass Sunstein's (2002) "Republic.com".

existing diversity to cross-check information from a variety of different sources in search for the truth in between. By using the example of blogs and describing them as mainly parasitic to traditional media Goldman adopts a simple sender-received model of communication as information transmission, a model reminiscent of Shannon & Weaver's mathematical theory of communication (Shannon 1948). The question of how and whether *new knowledge* gets created with the use of ICT is therefore again out of focus, although it should be a crucial topic given Goldman's interest in increasing the overall veristic value in a society. Despite my general acknowledgment of Goldman's early identification of these issues, I do think that his analyses of ICT suffer from the same narrow perspective which he also endorses for the topic of testimony. He considers both generic social practices, testimony and the use of ICT solely to be a means of knowledge transmission. However, as should have become obvious in my introduction of different social software applications, the Web has enabled an abundance of socio-technical epistemic practices which are generative of new knowledge and not only a means of signal transfer. To account for those practices a more comprehensive and broader understanding of the socio-epistemic functions of ICT has to be developed.

Don Fallis & the Epistemology of Mass Collaboration

Other than Goldman, it has been most notably Don Fallis, who related ICT and social epistemology. In an extensive chapter in *Annual Review of Information Science and Technology* (ARIST) he analyses and comments on the relationship between social epistemology and information science (Fallis 2006). In this paper, Fallis stresses the close relationship between information science and social epistemology. Libraries and information science more generally aims at helping people acquire knowledge by collecting, organizing and providing access to knowledge materialized in physical or digital media (Fallis 2006). Epistemology in general can help clarifying what knowledge *is* and social epistemology by focusing on the social factors and institutions that play a role in knowledge acquisition can enhance this perspective.

Fallis also notes that the term *social epistemology* was initially coined in the library sciences by Mary Egan and Jesse Shera, who argued for "[...] a new discipline [...] that will provide a framework for the effective investigation of the whole complex problem of the intellectual processes of society" (Egan and Shera 1952: 132, cited from Fallis 2006: 476). Information science according to Shera should focus on the "production, flow, integration, and consumption of all forms of communicated thought throughout the entire *social* fabric" (Shera 1970: 86, quoted from Fallis 2006: 482f), which is why a broader socio-epistemological perspective was required. According to Fallis (Fallis 2006) the connection between social epistemology and information science can and should be bi-directional: not only can information science be used for socio-epistemological research, e.g. by using bibliometrical data. Epistemology can also guide information science with respect to methodological issues.

If information science aims at providing knowledge, then clearly, social epistemology can be useful for a critical information science and Fallis lists several relevant issues (Fallis 2006: 502). One topic that should be of interest to information science is the philosophical analyses of testimony. Social epistemology might moreover help in weighing different epistemic objectives for information systems. Goldman's five standards for epistemic appraisal, i.e. power, speed, fecundity, reliability and efficiency may for instance be of particular interest (Goldman 1992: 195). Social epistemology might help in identifying different *types* of knowledge, in discerning different degrees of reliability that are needed in different situations. It might be of use for *knowledge organization*, e.g. for questions concerning classification and tagging. Finally, since Fallis acknowledges that epistemology and ethics are related in the analysis of ICT, he refers to some crucial value conflicts at this intersection: i.e. questions of intellectual freedom versus censorship, issues of privacy, intellectual property versus the freedom of sharing information, etc (Fallis 2006: 503ff).

Recently, Fallis also edited a special issue of the journal *Episteme* devoted to the "Epistemology of Mass Collaboration". This special issue is one of the first attempts to understand the role new media play for epistemic practices. Fallis argues that while people have for a long time collaborated for epistemic purposes in science and beyond, new technologies have enabled possibilities for mass collaborations on a much bigger scale. Some examples for such mass collaborations he refers to are Wikipedia, Yahoo! Answers and Digg.Com (Fallis 2009: 1). Given the ubiquity and importance of many of these social software applications for epistemic purpose, a critical epistemological analyses is indispensable and Fallis outlines some relevant questions:

> "How reliable are large collaborative projects that produce and disseminate information? What is the explanation for their reliability? Can large collaborative projects be reliable even if they do not make use of experts?

> Does the information produced by such projects count as testimony? Can we be justified in believing information produced by large collaborative projects? How should we go about deciding whether to believe information produced by such projects?"(Fallis 2009: 2)

Most of the papers in this journal focus on the example of Wikipedia and indeed Wikipedia is an interesting example for an epistemological analysis. First of all, it is a clear case of mass collaboration in a socio-technical system that serves epistemic purposes. How well it serves this purpose however, is hotly debated. While Wikipedia is rightly considered to be one of the great success stories of the Web , it also has many critics (e.g., Keen 2008, Sanger 2009, Waters 2007). Every once in a while, there are even political cries to ban Wikipedia and often it seems that some critics tend to throw the baby out with the bathwater. Instead of such clearly unrealistic overreactions, some more nuances analyses would be of great value. Several of the papers in this special issue deliver such analyses and this is why they are depicted below in some detail.

Wikipedia as a Source of Testimony

Deborah Tollefsen analyzes Wikipedia as a source of testimony. While most philosophical analyses of testimony have focused on testimonial transmission of knowledge between two individuals, Tollefsen has introduced the notion of *group testimony* (Tollefsen 2007). She argues that such a distinct concept of group testimony is necessary, because group testimony cannot be understood in a summative way. That means that group testimony is not just the sum of the individuals' testimony, i.e. testimony is not justified by the individuals of that group but by the group *itself* (Tollefsen 2009: 8).

Applying her notion of group testimony to Wikipedia, she asks what it is like to trust Wikipedia as a source of testimony (Tollefsen 2009). Are we - and if so to what extent – justified in believing Wikipedia as a source of testimony? Tollefsen affirms that Wikipedia can count as a source of testimony and that the question should rather be what exactly this *source* consists in. Analyzing the process in which Wikipedia entries are construed, she argues that Wikipedia consists of individual testimony and group testimony. *Mature* articles, articles which have been edited by different people and whose content is basically consented upon can be conceived as group testimony of the Wikipedia community. But since Wikipedia also includes many less elaborated articles, she compares Wikipedia to an "epistemic child" and argues that "[...] we should be careful [...] to monitor it closely for trustworthiness (Tollefsen 2009: 18).

Yet how exactly should we monitor Wikipedia? Tollefsen discusses a variety of reductionist possibilities to assess the trustworthiness of Wikipedia. She refers to Magnus (Magnus 2009)[110], who argues that common rules of thumb might not be suited to assess content on Wikipedia. Therefore, she argues that Adler's proposition to evaluate testimony by corroborating it with background beliefs rather than a monitoring of the speaker might be better suited to assess the trustworthiness of Wikipedia (Adler 1994, cited in Tollefsen 2009). Thus, instead of scrutinizing the content of Wikipedia for signs of trustworthiness, we might simply have to test to which extent the content of Wikipedia is corroborated by background beliefs we hold. Tollefsen concludes that either Wikipedia may evolve to become more mature so that we do not need to monitor its trustworthiness that closely anymore. The other possibility would be that the more familiar we get with the use of Wikipedia "[...] our learning mechanism [...] or epistemic sensibility [...] will develop in such a way as to be able to respond to group testimony in an unreflective yet critical way" (Tollefsen 2009: 22).

Trusting Wikipedia

P.D. Magnus asks whether and to what extent we should *trust* Wikipedia (Magnus 2009). He argues that although the average quality of Wikipedia articles is quite high, there is a huge variance between the qualities of different papers. Moreover, since articles can constantly change, one's assessment of the quality of the article can always only refer to its status at a very specific point of time. The epistemological problem given this variance and dynamicity then is how to distinguish the good from the bad and ugly, how to assess the quality of an article.

Magnus argues that several rules of thumb that we normally employ to assess the quality of epistemic content do not work properly in the Wikipedia environment. These strategies include the *assessment of authority*, i.e. does the information come from a reliable source; the *plausibility of style,* i.e. is the writing style adequate; *plausibility of content*, i.e. are there clearly implausible claims; *calibration*, i.e. is the majority of the other claims of this article correct; *sampling*, i.e. comparing claims from different sources (Magnus 2009: 79ff). Magnus argues that these strategies work differently and

[110] See below.

often less suited for Wikipedia than for other non-Web resources. Two examples are checking for implausible claims and cross-checking of content with other references. Magnus argues that since deleting inconsistencies and implausible claims is often the first step in editing a Wikipedia entry, those indicators of low quality are often deleted although the article is still as bad. Moreover, since many other websites link to Wikipedia or recycle content from it, the checking of independent sources becomes more difficult.

One aspect of Magnus analysis that I do very much appreciate concerns his emphasis on *empirical data about how users actually use Wikipedia* and his avoidance of frequently encountered simplifications as to whether Wikipedia should be used *at all* for epistemic purposes. As I have argued elsewhere (Simon 2009), I agree with Magnus that "[t]he question of *whether* we should trust *Wikipedia* becomes the question of *how* and to *what extent* we should trust *Wikipedia* (Magnus 2009: 77f). I do also agree with Magnus that we need to develop new methods and rules of thumb to assess the quality of online content. He argues that Wikipedia articles should be used as pointers to other resources, and that users should make more use of the history-pages of Wikipedia. Magnus acknowledges that many users might not take the time and effort to check these pages, although this may be epistemically beneficial. Fortunately, there is research on visualizing the history-pages in order to give some more intuitive and direct indicator of the reliability of Wikipedia articles. In Part 3 I introduce some of this research conducted by Ed Chi and his colleagues of the Socially Augmented Cognition Group at the Palo Alto Research Center (PARC) (e.g. Chi, Suh et al. 2008, Suh, Chi et al. 2008).

Magnus concludes his analysis by stating that "[...] teaching people to engage Wikipedia responsibly will require getting them to cultivate a healthy skepticism, to think of it differently than they think of traditional sources, and to learn to look beyond the current articles – and it will require learning to engage with it more responsibly ourselves" (Magnus 2009: 89). While I agree to almost all of these goals, I would insist that such healthy scepticism should also be retained when assessing traditional sources.

The Epistemic Culture of Wikipedia

Kay Brad Wray takes a different look at Wikipedia and compares the epistemic cultures of science and Wikipedia (Wray 2009). He argues that *collaboration* in those two fields functions quite differently. Not only is the stance towards knowledge itself different (creating new knowledge versus making it available), the knowledge producers (experts versus mainly amateurs) and the epistemic processes (democracy versus meritocracy) also differ between science and Wikipedia. Since the social structures, the norms and incentives as well as the goals of science and Wikipedia differ, Wray concludes that while an *invisible hand model of scientific rationality* can be assumed for science, this is not the case for Wikipedia (Wray 2009: 38).

He argues that due to these crucial differences between the types of epistemic collaboration, our reliance on the epistemic outcomes of these collaborative endeavors should differ as well. Wray concludes that Wikipedia is not a reliable source of knowledge and even if it were, it would be reliable for very different reasons that science is. I think that Wray's analysis of the differences between science and Wikipedia is a step in the right direction. Analyzing different socio-epistemic practices for their respective efficiency and reliability is certainly a major task for social epistemology. However, I am not quite convinced by some of his conclusions. First of all, he argues that Wikipedia cannot be a source of testimony, because we do not know anything about the person who has made a claim. To take ignorance of the testifier as a criterion not to accept testimony per se comes a bit surprising, given that Jonathan Adler for instance notes, that for some standard models of testimony in analytic philosophy it is actually assumed that the testifier is a stranger (Adler 2006). Moreover, referring to the notion of epistemic trust can shed a different light on this process of accepting rejecting Wikipedia as a source of testimony. More specifically, people may trust the *process* by which knowledge is provided in Wikipedia instead of the provider. I return to such a notion of procedural trust and its relevance on the Web in Part 3. A final point of critique that I see with respect to Wrays's analysis concerns the question of why the fact that Wikipedia functions *differently* from science *necessarily* implies that it functions *worse*. To my mind, it would be clever to suspend judgment for a moment and start analyzing on which occasions, which type of mass collaborations function best for which purposes. In Part 2 of this thesis, I target exactly this question.

Expertise and Wikipedia

In his article on "The Fate of Expertise after Wikipedia," Larry Sanger, defends the importance of experts for knowledge creation and for society more generally (Sanger 2009). Sanger founded Wikipedia together with Jimmy Wales, but left it in 2002. He has initiated a new Web-encyclopedia, Citizendium.org, which he considers to be superior to Wikipedia particularly due to its greater acknowledgement of experts. He starts his article with the observation of what he considers to be the paradox of Wikipedia, namely that "[...] Wikipedia is a striking popular success, and the quality of its articles, while uneven, is remarkably good; yet its success can be attributed in large part to the fact that it is both wide open and bottom-up" (Sanger 2009: 68). He asserts that while many Wikipedians would take a declaredly anti-expert stance, Wikipedia cannot serve as an example that experts are not needed anymore and that the wisdom of a crowd of amateurs can make up for expertise. Rather, it is the remainders of expertise in Wikipedia, such as giving reference to original work conducted by experts and the participation of experts that enables its success. If that reliance on experts were enhanced, Wikipedia would work even better, which is why, Citizendium.org offers exactly such an expert-friendly environment where knowledge can blossom. To my mind, Sanger confounds two issues that are frequently mixed up when the pros and cons of Web2.0 applications are discussed. One question concerns the relationship between experts and lay people, the other concerns the question of how many people participate in and epistemic endeavor. Although, most authors argue dichotomous either for single experts or masses of amateurs, I would argue that before making premature and simplified recommendations, one should devote some more time to analyze when and under which circumstances experts or amateurs, in solitude or united in different ways are best for which kinds of epistemic tasks.

The Epistemic Utility of Decision Markets

Only two authors in this *Episteme* special issue focus on examples of epistemic mass collaboration other than Wikipedia. George Bragues analyzes the relevance of *prediction markets* for a very special type of knowledge production: the forecast of the future (Bragues 2009). Bragues argues that for the longest time, certain individuals were chosen to predict the future. By now the role of the prophet or the augur is mostly taken over by scientists or other intellectuals. Nonetheless, the prediction is still mostly in the hands of individuals, or at least in those of small groups of experts. This situation changes with the emergence of prediction markets. Prediction markets are "[...] venues in which individuals trade securities whose value is tied to the outcome of a future event" (Bragues 2009: 93). While betting and gambling has existed for ages, the spread of prediction markets was greatly facilitated by the Internet and a variety of different types of prediction markets emerged. Bragues gives a brief survey of existing prediction markets and lists their characteristics, the opportunities and challenges that come with using them for epistemic purposes.

Bargues concludes his analysis by stating that prediction markets are a valid tool for forecasting and that they should be fostered. While some authors have interpreted the utility of decision markets and related tools as an argument to replace traditional experts with statistical procedures (e.g. Ayres 2007), Bragues delivers a more modest defense of the value of decision markets. He argues that decision markets should be regarded as *one tool among others* and that their value depends on circumstances. There are several prerequisites that have to be met for decision markets to function, such as *liquidity*, the ability to attract enough participants. Similar to James Surowiecki's prerequisites for a "wisdom of a crowds" (Surowiecki 2004)[111], Bragues also notes that a *diversity of opinions and cognitive strategies* are necessary for decision markets to function. Moreover, participants have to bet *independently* to avoid information cascades.[112] Finally, decision markets might not be suited for all types of questions, since they depend on *quantifiable outcomes*.

Bragues concludes his analysis of the epistemic value of decision markets by stating that while "[...] in non-empirical disciplines like literature and philosophy, it has long been customary to rely on individual geniuses to provide illumination, [...] prediction markets disclose that there is much knowledge to be gained by analyzing the fruits produced by the combined efforts of many people seeking to comprehend the same problems " (Bragues 2009: 103). I do appreciate Bragues thorough analysis a lot and also his modest and realistic assessment of the pros and cons of using decision markets for epistemic. In contrast to some more radical proposals, such as Ayres (Ayres 2007), his analyses strike me as a good starting point to assess socio-epistemic practices that are not based on deliberation and consensus, but on an aggregation of individual bets. This differentiation between different socio-epistemic mechanisms is explored in depth in Part 3 of this thesis.

[111] Surowiecki's account is portrayed in detail in chapter 8.
[112] On information cascades confer also Sunstein (2006) and Coady (2006).

Web2.0 versus Web3.0

Luciano Floridi finally assesses the success of Web 2.0 in comparison to the Semantic Web or Web3.0 (Floridi 2009). He argues that the reason why Web2.0 is so successful as compared to Semantic Web projects or earlier approaches in Artificial Intelligence is because Web2.0 exploits existing intelligence smartly instead of trying to construct new forms of intelligence. So far, he argues "[...] humans are the only semantic engines available" (Floridi 2009: 32) and any attempts to automate semantics in the Semantic Web are still based on human input. As a result, Semantic Web applications are either exciting science fiction (when "semantic" in Semantic Web is taken seriously) or realistic trivialities [...]" (Floridi 2009: 26), which do not go beyond metasyntax.

Web2.0 works very differently. By exploiting human intelligence in an aggregated and collaborative way, Web2.0 applications have been highly successful in providing and filtering information. One example that Floridi uses to explain the difference between Web2.0 and Web3.0 are folksonomies. Folksonomies are the result of the aggregation of bottom-up, user-generated tags to organize content for later retrieval. They provide the opposite to top-down classifications, to expert-generated ontologies. This is not the space to discuss the pros and cons of classification versus tagging, but for the moment it suffices to note that although folksonomies might come with some problems, they are much more robust and adaptive than ontologies. Floridi summarizes his comparison by stating that because of these differences "[...] Semantic Web is a well-defined mistake, whereas the Web 2.0 is an ill-defined success" (Floridi 2009: 33), because Web3.0 does not exploit the collective intelligence of humans in the way Web2.0 applications do.

Interestingly a similar trend can be observed when taking a look into the history of learning software (Issing and Klimsa 1997). A lot of work was spent on building *intelligent tutorial systems* (ITS), which should adapt to users' needs by analyzing their learning progress and modifying the teaching material accordingly. Despite this energy, the systems never quite met the expectations of the developers or the users. The learning models implemented in the systems were too much simplified to adapt appropriately to the users' needs. The real change by contrast was brought about by hypertextual environments. The adaption in these systems was much more effective and efficient, because it was now the users who could adapt their own navigation through the system to their needs. Although successful navigation was not always ensured and required a certain amount of expertise and hypertext literacy from the user, such hypertextual systems took their advantage from the fact that they used existing – human - intelligence instead of trying to simulate it technically.[113] To my mind a crucial lesson that can be learned from these failures is that one should take a great deal of effort in *distinguishing which epistemic practices should be done by humans and which should better be delegated to the computer*. Since the respective intelligence of human and computers differs, it would be stupid not to take the best from both worlds and to combine it fruitfully.

6.4.4 Socio-Epistemological Analyses of ICT: Conclusions & Critique

I consider all the before-mentioned analyses to be interesting starting point for assessing the socio-epistemic relevance of social software. They highlight various aspects that any thorough and comprehensive social epistemology of epistemic social software has to address. However, I think that these are only the first beginnings and there still is a lot of work to be done in analyzing the new possibilities, the new risks and chances that come with the emergence of new socio-technical networks.

First of all, there are several shortcomings within social epistemology as portrayed in the last chapters, which have also hampered the socio-epistemological analyses of ICT. As was shown for Alvin Goldman, most of his analyses are clearly based on a simple sender-receiver model of testimonial knowledge transmission (Goldman 2003, Goldman 2008). While transmission of existing knowledge might be one crucial goal that ICT serve, it clearly is not the only one. Therefore, it will be essential to broaden the frame and assess the role different information and communication technologies play for the *creation* of knowledge. Moreover, I consider the models of information and communication technologies to be too simplistic. As I have argued in Part 1, epistemic social software should better be understood as socio-technical epistemic systems consisting of multiple agents, and not as mere tools or technologies in the hands of rational human agents. I think this narrow focus on social software as information and communication *technologies* and the neglect of the entangled nature of the social, the technical and the epistemic is also one of the reasons for the narrow focus on knowledge transmission.

[113] For further information on learning software cf. for instance (Issing and Klimsa 1997).

Secondly, we need to analyze *more and different examples of epistemic social software*. To understand the relevance of socio-epistemic practices on the Web, we have to assess applications beyond Wikipedia and decision markets, such as recommender software or social bookmarking systems or search engines (cf. Mager 2009, Röhle 2009). I have focused on recommender systems and social bookmarking systems before (Simon 2007; Simon 2008; Simon 2009), and will address these and other examples in the remaining chapters of this thesis as well.

Thirdly, we also have to broaden the *kinds of questions* we ask. Some of them will have to deal with the different types of people involved in such socio-technical systems. For instance, to what extent are different user groups affected differently by socio-epistemic practices on the Web? Who gets included and who gets excluded by which mechanism? What are the possible epistemic effects of such differential treatment? To what extent are those epistemic effects related to ethical and political considerations? We may not only have to ask more and different epistemological questions about the relevance of ICT for knowledge, but also have to investigate the relationship between epistemology, ethics and politics.

7 Summary and Conclusions of PART 2

Before developing my own comprehensive socio-epistemological framework to analyze epistemic social software, I briefly summarize my position on the issues raised in this second part of the thesis along the major debates that I have outlined.

Testimony

A broader notion of testimony that includes its relevance for the creation of knowledge is needed for a full account of testimony. To understand processes of knowledge distribution on the Web, however, even narrow conceptions of testimony may be helpful.

Trust

Epistemic trust is a crucial concept for any socio-epistemological analysis of epistemic social software. Trust in agents, processes and trust in knowledge as content are indispensable prerequisites not only for the reception of testimony, but also for knowledge creation. However, this trust is hardly ever blind - and it should not be blind either. Assessing the epistemic trustworthiness of epistemic agents and processes is crucial for being a responsible knower. The assessment of epistemic trustworthiness is a judgment based on criteria which can be more or less valid. If they are invalid, not only can error, i.e. epistemic mistakes, be a result, epistemic injustices may take place as well.

Distribution of Epistemic Labor

I prefer the term *epistemic* labor over the more widely used term cognitive labor because it avoids implicit assumptions that epistemic practices may be only mental. Rather epistemic labor is meant to include all sorts of practices that are related to the creation or distribution of knowledge, i.e. handling machines, manipulating artifacts, using mathematical procedures, gathering data through various methods, etc. Further, although the term *distribution of cognitive labor* usually refers to distributions of research effort within scientific communities as a whole, I will use the term more broadly. I argue that distribution of labor is a crucial *starting mechanism* for any socio-epistemic process involving multiple agents.

Consensus Formation

Consensus formation is a central topic in social epistemology. I consider consensus formation to be a possible mechanism of *closing* socio-epistemic processes, but it is not the only one. My own socio-epistemological framework indeed is based on a tripartite classification of closure mechanisms. Further, I agree that with Solomon and Longino that consensus on methods, theories or approaches in science should be avoided as long as possible and that epistemic pluralism should be supported. Nonetheless, in cases where consensus is needed, normative methods of achieving it may be useful. I consider Keith Lehrer's model of rational consensus to be of particular relevance for two reasons. Not only does he provide a rational model for consensus, he also delivers insights concerning the use of reputation for epistemic purpose. The pros and cons of using evaluative social information, i.e. reputation, to weigh epistemic statements or epistemic agents are highly relevant for my analyses in the next chapters. While the benefits of using reputation as an epistemic proxy are acknowledged, the dangers of epistemic injustice have to be kept in mind.

Complementarity

I side with requests for more radical conceptions of social epistemology that are not merely an addendum to individualistic epistemologies, but question the foundations of individualistic approaches. In particular, I agree that knowledge is a social status that fundamentally depends on communities. While individuals may believe, see, infer, knowledge only comes into existence though communal vetting of content. This does not imply however, that insights obtained from complementary social epistemologies may not be suited to analyze certain *socio-epistemic aspects* of socio-technical epistemic systems.

What's Knowledge and Who is to Decide?

I adopt Longino's tripartite notion of knowledge for this thesis (Longino 2002c). Her distinction between *knowledge as content*, knowledge as a set of *knowledge-productive practices* and from *knowledge as cognitive agency*, i.e. as a state of a person ('knowing') provides a solid foundation for analyzing socio-epistemic practices in socio-technical epistemic systems. I also share Martin Kusch's and Helen Longino's receptive stance towards the social studies of science. As should have become obvious from my analyses in Chapter 3, I consider the insights obtained in these fields to be highly relevant for the development of any social epistemology that aims at being not only normatively appropriate, but empirically adequate. Indeed, I consider empirical adequacy to be a prerequisite for normative adequacy.

The Utility of Truth

There are several reasons why I consider Longino's notion of *conformation* to be the most attractive framework for analyzing socio-epistemic practices and products on the Web. First of all, the concept of conformation can be applied to non-propositional content and to epistemic *practices*. Moreover, conformation as opposed to truth comes in degrees and it can be assessed for different dimensions depending on the purpose of epistemic tasks.

The Relevance of Epistemic Communities: Contextualism and Relativism

I agree with Martin Kusch and Helen Longino that epistemic standards are always local and that declaring epistemic content to be knowledge depends on negotiations based on such local standards. I further share Longino's acknowledgement of shared standards which evolve over time, but are nonetheless temporarily binding. However, I do not think that such a contextualist position can avoid being relativist. Here I side with Martin Kusch: if relativism denotes that truth and falsity depend on a communities interests, exemplars and goals but does not imply that all statements are only "relatively true" or that one can make a theory true by simply wanting it to be true, then the position taken in this thesis is relativistic.

Methodological Issues: Naturalism and Quantification

I share the commitment of many social epistemologists naturalism understood as the receptiveness to insights from empirical sciences. Yet I do not consider knowledge to be a *natural kind*. Moreover, many quantitative methods, especially various stochastical techniques are and should be important for social epistemology. Indeed, in Chapter 10 I outline a type of epistemic sociality that is entirely based on statistical aggregation of data and defend it as an important socio-epistemic mechanism. However, there are also inherent dangers of quantifying the qualitative and those have to be monitored. Issues of particular concern are flaws in the initial quantification of the qualitative, biases, and the danger of misusing the power of numbers.

Normativity

I agree with most social epistemologists that social epistemologies cannot remain purely descriptive but that a normative stance is crucial. Normativity is necessary for critical analyses just as much as for suggestions on how to improve socio-epistemic practices and systems. In either case, numerous criteria for assessment are available and they need to be chosen and weighted depending on the context and the goals of epistemic inquiry.

Towards a Social Epistemology for Social Software

The initial approaches to analyze ICT and social software from a socio-epistemological perspective portrayed in this chapter are interesting, but they can only be considered starting points for a more comprehensive analysis of socio-technical epistemic systems. Many approaches are hampered by a narrow focus on knowledge distribution. Moreover, more and different examples of social software should be analyzed, such as recommender systems, search engines, social bookmarking systems. Finally there are more and different questions to be asked. A comprehensive framework for analyzing epistemic social software must enable addressing all or at least most of the issues raised in the previous two chapters. I have now outlined my stance towards the most central debates in social epistemology. My own socio-epistemological framework to be developed in the next chapters is based on these background assumptions.

PART 3

8 Towards A Socio-Epistemological Framework for Epistemic Social Software and Beyond

8.1 The Fundaments

In the following chapters I outline a new socio-epistemological framework to analyze epistemic social software, i.e. social software whose *primary* purpose lies in the creation, dissemination or evaluation of knowledge. In emphasizing the processes of collective knowledge *creation*, this framework goes beyond previous socio-epistemological analyses of ICT which mainly focus on the *distribution* of knowledge only.

The key commonality of epistemic social software is that multiple epistemic agents interact with each other, with technological artifacts and infrastructures, with objects, i.e. with the world around them more broadly conceived, to create epistemic content. Based on insights from the field of STS I argue that social software should better be understood as socio-technical epistemic systems in which multiple human and non-human agents interact. I ascribe agency to non-human or non-animated agents to stress their central role in processes of knowledge creation. Knowledge is the result of entangled processes taking place within such socio-technical networks comprising of different agents – human and non-human - who fulfil different epistemic tasks. Indeed one of the crucial insights from my analyses is that in certain socio-technical epistemic constellations, technologies - non-human epistemic agents- are ascribed more *authority* than human epistemic agents. In those cases, *trust* is placed in algorithms rather than in human knowers. Hence, applying the term epistemic agents to human as well as non-human agents is meant to indicate a dissolution of boundaries between human and technologies with respect to their agency, their roles within epistemic processes.

Ascribing *agency* to artifacts has to be distinguished from ascribing *intentionality* to them. The attribution of authority to *algorithms* for instance does not imply the attribution of intentionality to such algorithms, but it stresses their active role within epistemic processes. It emphasizes that they follow certain dynamics; that they have certain effects; that they have a life on their own instead of being mere tools in the hands of fully rational and intentional agents. This active role has to be acknowledged and addressed in any analysis of socio-technical epistemic systems. This ascription of agency to technological artifacts also does by no means imply that I wish to equalize humans, technological artifacts, animals, unanimated objects etc. in all respects. As argued in Chapter 3, there are epistemological as well as ethical reasons for not completely equalizing human and non-human epistemic agents. For this reason I frequently distinguish between human and non-human epistemic agents in the following chapters. While intentionality may be one important difference, the problems around the attribution of *accountability* and *responsibility* are of even greater concern. Both aspects have been emphasized in particular by feminist STS scholars in their ethical-epistemological analyses of technologies. Accountability and responsibility have twofold meanings for critical analyses of ICT. Not only do we want to attribute responsibility and accountability to *other* agents within systems, especially in cases of failure or error. We also want to assume accountability and responsibility for our *own actions* within socio-technical epistemic systems to be responsible and accountable users – and knowers. Whether it is possible to be responsible for one's own actions, however, depends on the extent to which one understands one's own actions within such systems. This in turn requires a certain amount of *transparency* – a criterion for which systems designers are responsible. Hence, a) there are differences between human and non-human epistemic agents with respect to questions of responsibility and accountability and b) the possibility for participants in such systems to be responsible for their actions depends upon criteria of the system, such as transparency, for which the system designers are responsible and to be held accountable.

A crucial characteristic of social software, which distinguishes it from ICT more broadly conceived, is that human agents do not only interact with technologies, but also with other human agents in various forms. This interaction is technically mediated in various ways and to varying degrees, but nonetheless the participation of *multiple human agents* is a crucial feature of social software. Within *epistemic* social software applications, which are at the heart of this thesis, such agents participate for *epistemic* purposes.[114] If

[114] This does not imply that the only motivation for participating in such systems is epistemic, i.e. related to the creation or distribution of knowledge. Indeed, as Yochai Benkler (2006) argues, for commons-based

people interact for epistemic purposes, they have to communicate epistemic content. That is, they make use of existing epistemic content, process it and share their own results. From this procedural perspective the focus of social epistemology on the distribution of existing knowledge makes some sense: since knowledge-producing practices depend on previous knowledge, the distribution of existing knowledge is an important function of ICT. However, distributing knowledge is clearly not the only epistemic process within social software. At least as important are processes of *knowledge creation,* yet this aspect has so far been rather neglected in socio-epistemological analyses.

Talking about knowledge distribution and knowledge creation presupposes a definition of what is understood by knowledge. As outlined in Part 2, my conception of knowledge is strongly influenced by the socio-epistemological theories developed by Martin Kusch (Kusch 2002) and Helen Longino (Longino 2002c). With them I conceive knowledge to be a social status that can be ascribed to epistemic content by a community. Knowledge further is a success term labelling epistemic content that has survived critical scrutiny from multiple agents and satisfies communal standards. Knowledge is the result of socio-epistemic processes, in which interdependent and situated epistemic agents interact in different ways and to different degrees. Moreover, I adopt Longino's tripartite classification of knowledge into *knowledge as content*, *knowledge as cognitive agency* and *knowledge as knowledge-productive practices*, because it helps avoiding frequent misunderstandings concerning the term knowledge.

Longino's tripartite notion of knowledge has been portrayed in detail in Chapter 5. To recapitulate, she argues that many misunderstandings concerning knowledge are a result of the fact that the word "knowledge" is used to describe quite different things. *Knowledge-productive practices* describe the processes involved in transforming various inputs into representational, epistemic outputs. *Knowledge as cognitive agency* describes the three-term relation between a subject, an object and a representation. *Knowledge as content* finally refers to the body of knowledge, to what is known in its materialized form. Socializing the three modalities of knowledge made Longino conclude that knowledge-productive practices and their modes of justification vary in different

---

peer production, the motivations to take part in the collective production of content on the Web may comprise of different combinations of epistemic and non-epistemic, altruistic and egoistic motives. Nonetheless, the systems of concern are characterized by a clear epistemic dimension even if the motivations for participation may be diverse.

contexts; cognitive agents are interdependent and situated; and knowledge as content is plural and depends upon effective transformative criticism by differently situated epistemic agents who nonetheless share epistemic standards (Longino 2002c: 122).

Adopting such a perspective on knowledge has consequences for the analysis of epistemic social software. Such software as socio-technical epistemic systems consists of situated and interdependent epistemic agents that engage in knowledge-productive practices by interacting with one another, with their objects of inquiry and the technological infrastructure connecting them. To reach their epistemic goal, agents make use of existing epistemic content, ideally *successful* epistemic content, i.e. knowledge as content that has survived critical scrutiny. They use this content as input for knowledge-productive practices through which new epistemic content is produced that can then be subjected to public scrutiny and may end up being knowledge if this status is ascribed to it by a community.

All these interactions are based on *trust*: trust in other epistemic agents, trust in epistemic content that is used as input, trust in artifacts, and trust in epistemic processes themselves. Trust is a fundamental ingredient of epistemic processes, without trust there can be no knowledge. Yet as I argue in detail in the next chapters the *loci of trust* differ profoundly among different socio-epistemic systems.

Another crucial consequence of distinguishing knowledge-productive practices and knowledge as content concerns the acknowledgement of the *temporal structure of epistemic processes*. If knowledge as content serves as input and as output of epistemic processes, knowledge as content depends upon a process of closure, upon the (temporary) termination of epistemic processes. In order to declare epistemic content to be knowledge it must have been made subject to scrutiny. Hence content has to be put into a *format* and it has to be made *public*. Epistemic processes are not unstructured processes, which continue in eternity. Even if they continued in eternity they are partitioned by temporary terminations, closures of processes, which result in closed, presentable epistemic products. These closed and presentable products are the pivots, which serve as reference points for socio-epistemic practices. *Knowledge as content* then refers to epistemic products resulting from temporary closure of socio-epistemic processes.

In the following chapters I outline a framework to analyze epistemic social software that is based on a *tripartite classification of different socio-epistemic mechanisms of closure* employed in such systems. That is, I propose a framework to analyze socio-technical epistemic systems that unfurls epistemic processes from its end, from the mechanisms of termination. This closure puts an end to a socio-epistemic process by creating the *collective epistemic result* of this process. In that sense creation of epistemic products depends upon the temporary termination of socio-epistemic processes. This end, of course, and the product that is produced by it, then offer the venture point for another epistemic process to set it. The epistemic product that is generated can be used as input for further epistemic processes. This is what I mean by saying that knowledge is input and output of epistemic processes.

Indeed, to close socio-epistemic processes actually two processes need to be closed: the *individual epistemic processes* in which individual epistemic agents are involved and the *social or collective epistemic process* comprising of these individual processes. In each socio-epistemic process in which multiple epistemic agents are involved and aim at certain epistemic goals the following steps have to be taken. First, each individual has to provide his or her epistemic result - as minuscule as it may be. Given the generality of my framework as well as the generic yet nuanced understanding of knowledge adopted from Helen Longino, the range of such individual epistemic contributions is immense: it can be anything from a judgment of a person on the quality of a book to a written book itself, from the design of experimental stimuli to experimental result, from a diagram to a PhD-thesis on the use of diagrams for philosophical reasoning. In either way, for this individual epistemic result to be considered knowledge or even to be used for further knowledge creation, it has to be externalized. The process of its production has to be terminated to produce a sharable and addressable output. However, to obtain a *collective result*, the socio-epistemic process needs to be closed as well. The individual epistemic products need to be put together in one way or another; a decision needs to be made about how to construe the collective result from the individual results. These processes of closure form the core of my socio-epistemological framework. More specifically, I argue that three generic mechanisms of closing socio-epistemic processes exist: *integration, aggregation* and *selection*. In the following chapters I depict their characteristics and the differences between them in some detail.

8.2 Two Additional Sources of Inspiration

Before I start outlining my model in more detail, I have to introduce two more approaches, which have served as important input for the specification of these closure mechanisms: Yochai Benkler's work on commons-based peer production (Benkler 2002; Benkler 2006; Benkler and Nissenbaum 2006) as well as James Surowiecki's introduction of the term "wisdom of the crowds" (Surowiecki 2004). This further hiatus before presenting my own model needs to be explained.

To a certain degree, the works of Surowiecki and Benkler do not fit into the previous chapters, because they are neither part of the socio-epistemological nor the STS discourse. Nonetheless, they address socio-epistemic topics and, albeit in different ways and to different degrees, epistemic social software. Benkler's work on the transformative power of the Web is characterized by an analysis of different social software applications from an economic and legal perspective and methodologically based on *transaction cost analysis*. Thus, our perspectives as well as our theoretical and methodological frameworks differ. Yet we analyze the same phenomena and ask similar questions. In particular his work delivers crucial insights for socio-epistemological analyses of epistemic social software for two reasons. First, his notion of commons-based peer production can shed light on important aspects of socio-technical epistemic systems that employ *integrative mechanisms of closure*. Hence, I refer to Benkler in particular for the characterization of my first type of epistemic sociality: Epistemic Sociality:Integration (ES^I). Second, Benkler exposes important characteristics of socio-technical epistemic systems and processes on the Web which can serve as *criteria* for the analyses of epistemic social software employing different mechanisms of closure, i.e. integrative, aggregational or selective mechanisms. Hence, I introduce Benkler in some detail, because I make use of his insights to a) characterize one specific type of epistemic sociality and b) to deliver criteria for the analysis of systems employing all types of epistemic sociality.

James Surowiecki's work is also characterized by an economic perspective, but it is less scholarly than Benkler's. Indeed, Surowiecki's book "The Wisdom of Crowds: why the many are smarter than the few and how collective wisdom shapes business, economies, societies, and nations" has been a bestseller and the notion of the *wisdom of the crowds* is used extensively in the discourse around the merits of the Web2.0. The reason for this is less the relevance of Surowiecki's own analyses of social software, but rather the fact that the basic mechanisms of his *wisdom of the crowds* are characteristic of many Web2.0 applications. I argue that Surowiecki's book has been so influential, because he addresses a new type of epistemic sociality, a new mechanisms of orchestrating epistemic labor, namely the collective creation of epistemic content through *aggregation.* Hence, I use Surowiecki's work to characterize another type of epistemic sociality: Epistemic Sociality:Aggregation (ES^{A}).

Despite numerous differences between my approach and these two approaches, Benkler's and Surowiecki's work has crucially informed my analyses because we ask similar questions concerning the coordination of epistemic processes in which multiple epistemic agents are involved. Therefore I make use of their insights to amend my own socio-epistemological framework

8.2.1 Yochai Benkler's Commons-Based Peer Production

One of the most interesting accounts of how changes in information and communication technologies affect the ways in which we produce intellectual content was delivered by Yochai Benkler.[115] In his book "The Wealth of Networks: How Social Production Transforms Markets and Freedom" Benkler delivers an analysis of the impact of ICT on production processes in our contemporary society (Benkler 2006). While his analyses are based on economic theory, the conclusions he draws target legal and political action. With reference to *transaction cost analysis*, Benkler assesses and explains the emergence of a third mode of productions, in addition to *markets* and *firms*. Benkler labels this third mode *commons-based peer production* (CBPP). He argues that while sharing and exchange have always existed as modes of production in societies, it is due to the availability and ubiquity of personal computers connected to the Internet that a mode of production emerged which operates on the *same scale* as markets and firms but on entirely *different principles*. Moreover, commons-based peer production has *systematic advantages* over markets and firms in the production of informational goods, i.e. in the production of knowledge, information and culture.

Although the emergence and success of numerous examples of CBPP might speak for itself, it is not at all clear that the hopes attached to it will come true and that the full

[115] Unless otherwise noted, my portrayal of his notion of commons-based peer production in this section is primarily based on Benkler (2006, especially Chapter 3 & 4) and Benkler (2002).

potential of CBPP can be tapped. Benkler argues that we are currently at a moment of transformation in which we have to take a stance and in which the political and legal decisions we take will be highly consequential. He urges that we have to decide upon the direction in which our society quite generally should be heading. In particular, Benkler argues that we have to fight strict copyright and intellectual property rights and promote nonproprietary, commons-based information production. By combining empirical data from the social sciences with economic reflection, Benkler shows that strict enforcement of intellectual property rights, limiting access to intellectual products is detrimental to creativity and innovation. In brief, he argues that information as opposed to other market goods has two key characteristics. First of all, it is *non-rival*, i.e. the value of information is not diminished by use. If I share my knowledge with you, I do not have less knowledge afterwards, nor is the value of my knowledge decreased.[116] Secondly, information is not only the *output* of a production process, but it is also its *input*. To use socio-epistemological terminology, knowledge as content serves as input for knowledge-productive processes yielding in new epistemic content which may end up being acknowledged as knowledge (Longino 2002c). In order to produce new knowledge, people need access to prior intellectual products, to existing knowledge and information. If this access to existing knowledge is restricted, decreases in innovation and knowledge creation are the result.

Accordingly, in order to improve creativity and the production of knowledge, access to intellectual goods has to be free and not restricted by property laws that only benefit their few owners. Benkler thus contravenes positions according to which property rights are needed as incentives, because people would stop producing content if they could not hold exclusive rights over them. He refutes these claims arguing that many highly knowledge intense sectors, such as science, are structured around non-exclusive and/ or non-market strategies. Given that the positive effects of strict intellectual property rights are minor or even altogether questionable compared to the negative effects, he argues against such strong intellectual property rights. Realizing that especially mass media monopolists have a lot to lose in this battle, Benkler considers his work to provide not only an empirical or conceptual, but also a *moral framework* for an argument to support CBPP legally and politically (Benkler 2006: 472). The scope of changes that rest upon

[116] This is at least the case if we ignore for the moment those instances in which exclusivity might have an additional, possibly financial value for me, e.g. in journalism.

the seemingly minor decisions about intellectual property rights is enormous. The decisions that are made will affect major questions of justice, freedom and human welfare. Benkler delineates the effects of strong intellectual property rights for a huge variety of issues, ranging from political freedom and democratic participation over cultural capital to health care and agricultural infrastructure in developing countries (Benkler 2006).

To my mind many aspects of his analyses do have socio-epistemic relevance and there are many important insights to be gained from Benkler's work despite his focus on economic theory and political-legal interventions. First and foremost, his model of CBPP delivers a clear case of socio-technical epistemic processes on the Web2.0. Not only does CBPP refer to the organization of multiple human agents interacting with each other via technological infrastructures. Such agents are also engaged in the collective production of *intellectual goods*, the production of information, culture and knowledge. To my mind, two clusters of problems addressed by Benkler are crucial for any socio-epistemological analysis of socio-technical epistemic systems and should hence be considered when developing socio-epistemological frameworks to analyze such systems. The first concerns the *different epistemic functions of communication for peer production*. The three functions which Benkler distinguishes are *uttering content, relevance/accreditation* and *distribution* (Benkler 2006: 68ff). I relate Benkler's analysis of the collective realization of these three functions on the Web to socio-epistemological considerations on different *phases* in the process of knowledge creation. The second cluster concerns the organization of collective action. It comprises of answers to the questions of a) how to allocate tasks to agents, b) how to partition and re-integrate epistemic tasks and c) how to motivate participants to engage in collective action.

Three Epistemic Functions of Communication: Uttering, Accrediting and Distributing Content

Benkler describes three different functions of communication in the peer production of information, knowledge, and culture more generally: uttering content, relevance/accreditation and distribution (Benkler 2006: 68ff). These three functions have their counterparts in knowledge-producing practices and have therefore been addressed – albeit with different terminology – in socio-epistemological theories. First of all, it seems plausible to draw a parallel between Benkler's two functions of *uttering* *content* and *relevance/accreditation* and Helen Longino's two aspects of reasoning, i.e. constructive and the justificatory (Longino 2002c: 103). As noted before, while individuals can create content, the creation of *knowledge* depends on a communal accreditation of content, on the attribution of knowledge as a label for successful epistemic content that has survived critical scrutiny. Thus, by analyzing the distribution of this accrediting process over multiple agents on the Web, Benkler clearly highlights a process of high socio-epistemological relevance. Finally, the *distribution* of existing knowledge has not only been the aspect under which ICT has mostly been analyzed in social epistemology. From the perspective of knowledge creation, this distribution of knowledge as the output of previous socio-epistemic processes is important in so far as is makes this content available as input for new socio-epistemic processes by which new epistemic content is produced that finally may end up being attributed the status of knowledge by a community as well.

The Web offers a platform where everyone with the necessary technical equipment and skills can produce content and make it publicly available. Hence, the Web itself is an example of commons-based peer production. On the Web people provide content for one another – often for motives other than financial ones, content is continuously uttered. In CBPP collective, distributed action, however, is not only relevant for the *creation* of content, but also for the *assessment of relevance and for accreditation*. That is, not only is content produced collectively, the assessment of the relevance and the control for quality is often also done collectively on the Web. With respect to social epistemology, this difference between the initial uttering of content and the assessment of relevance can be compared to the stages in which knowledge gets created. First some intellectual content has to be produced, but only after a communal evaluation it may be considered knowledge (e.g. Kusch 2002, Longino 2002c). Clearly not all processes of relevance assessment necessarily lead to knowledge claims being justified. Nonetheless, the opposite holds true, without evaluative assessment, knowledge does not get created.

On the Web, relevance assessment and accreditation comes in more or less automated forms. Prime examples of automated processes are recommender systems, such as the ones employed by Amazon.com and search algorithms like Google's PageRank. Examples of less automated processes can be found in the Open Directory Project

(ODT).[117] The ODT is a human-edited Web-directory, which aims at organizing the content of the Web by topics. Volunteers decide upon the acceptance of links to this comprehensive directory. However, this human-centered form of knowledge organization lost much of its relevance in the last years. While in 2004 the directory was still on the main page on Google, it was degraded gradually into the submenu until it finally disappeared from the options on Google in 2008 (Rogers 2009).

Clearly, simply ranking content or providing a classification for it does not make it *true*. However, as Helen Longino (Longino 2002c) and Paul Thagard (Thagard 1997a) have argued, truth may not be the best concept for assessing scientific theories and practices. What also matters is their fertility, the extent to which they lead to new research and to empirical results. A similar stance might be taken when analyzing (automated) forms of accrediting and relevance assessment on the Web. If the information I get matches my epistemic needs, this mechanism for filtering for relevance is epistemically successful. Thus, accreditation and relevance are not knowledge itself, but they are *instrumental for knowledge*. They should and actually cannot be assessed with a dichotomous true-false distinction, but rather with respect to their fertility, the efficiency, etc. Alvin Goldman, for instance, has presented five standards to assess epistemic practices: reliability, power, fecundity, speed and efficiency (Goldman 1992: 195), which Paul Thagard has already applied fruitfully to ICT (Thagard 1997b). I come back to the usefulness of these standards below, but for the moment I only assert that these standards may well be suited to assess different mechanisms of accreditation and relevance assessment on the Web. If these mechanisms deliver information needed to create new knowledge they are epistemically beneficial. If they deliver information that inhibits the creation of knowledge or leads to more ignorance and error, they are epistemically detrimental.

Benkler delivers numerous examples of how content gets uttered, accredited and distributed on the Web to support the idea that commons-based peer production indeed is a mode of production that competes with markets and firms and may even overtrump them when it comes to the creation of intellectual products (Benkler 2006). In the following I list some of the examples that Benkler refers to in order to show the ways in which multiple agents are engaged in the uttering, assessment and distribution of intellectual content.

[117] http://www.dmoz.org/ [date of access: 13.02.2010]

A particularly prominent example of collaborative content creation on the Web is *Wikipedia.*[118] In Wikipedia thousands of volunteers from all over the world collaborate to create a comprehensive, multi-lingual encyclopedia for free. The accreditation and quality control mechanisms here are a mix of technical features (e.g. history page, ease of revision to former status), social norms (such as Wikipedia's Neutral Point of View policy), and a hierarchical division in users and administrators with special rights and obligations (Benkler 2006: 70ff). In an earlier paper, Benkler also refers to Kuro5hin.org[119], a platform where essays on the interplay between technology, media, politics and culture are published and commented on. The quality control mechanisms here are similar to academic peer review but get combined with post-publications peer comments (Benkler 2002).

Another example in which relevance assessment comes closer to forms of quality control known from science is Slashdot.org. Slashdot.org is a platform, which according to their self-description provides "News For Nerds. Stuff That Matters"[120]. The "stuff" that they provide is primarily information on technology-related topics, but the process by which this content gets filtered is quite complex and epistemically interesting. Users can submit content that they consider interesting and want to share. The filtering of these submissions lies in the hands of paid employees, who basically function as editors. However, after this filtering, the published content can be commented upon and evaluated by the users again. These comments on papers are subject to scrutiny as well and can get rated by "moderators". The role of the moderator is temporarily ascribed to users, who fulfill certain criteria: they have to be logged-in and not anonymous, they must have participated for a while and their comments must have been evaluated positively by others. Once they have obtained this status, they can rate five comments within the next three days. Their review of the comments is assessed as well by a process of meta-moderation of yet another group of moderators. Thus, slashdot.org provides a peer review mechanism that is amended with several checks and balances. They combine hierarchical structures with egalitarian means to acquire these statuses temporarily. Benkler concludes that Slashdot.org with its complex review mechanisms provides an example of how accreditation and relevance assessment can be done in a distributed way instead of having to rely on professional accreditation experts.

[118] http://en.wikipedia.org/wiki/Wikipedia [date of access: 13.03.2010]
[119] http://www.kuro5hin.org/ [date of access: 13.03.2010]
[120] http://slashdot.org/ [date of access: 05.11.2009].

Another example that Benkler introduces to show how content is produced via commons-based peer production is related to science as well: *NASA clickworkers.*[121] In this project, volunteers can mark and classify craters in maps of the Mars (Benkler 2002; Benkler 2006). Formerly, such tasks had to be done by experts. In Clickworkers, they are split up over thousands of lay people who perform these highly modularized tasks of marking and classifying at the instance of a click. No particular expertise is needed in the performance of these single tasks. Expertise is needed however in the coordination and integration of these tasks and in the provision of control mechanisms to weed out error and two NASA employees were in charge of these professional tasks. In particular, the control mechanisms employed to minimize error are based on redundancy and automated averaging of the input.

Many social epistemologists have focused in particular on the role that ICT plays for the distribution of knowledge (e.g. Goldman 2003, Magnus 2009, Tollefsen 2009). Benkler (Benkler 2002; Benkler 2006) agrees that the benefits of the Web and the Internet with respect to the ease, speed and low costs of distributing information are enormous and cannot be underrated. If knowledge is needed as input for (socio)-epistemic processes, then the distribution of knowledge is of clear socio-epistemic relevance. Moreover, in certain Web projects, distribution is amended by value-adding practices. One example is the *Project Gutenberg*, which provides books as well as more recently also audio books, whose copyrights have expired. Yet Project Gutenberg does not merely distribute knowledge. Instead it is value-added distribution, because volunteers transcribe the books into electronic formats and by doing this make them available electronically. Moreover, in the course of this process, *proofreading* is an essential mechanism of quality control and this task is done by volunteers as well. On the site *Distributed Proofreaders*[122], users can proofread single pages and by this help improve the quality of the electronic versions of the books.

Having introduced these examples of in which people collaborate on the Web, Benkler draws his conclusions on this phenomenon he labels CBPP. He describes *commons-based peer production* as a new mode of production that is "[...] "radically decentralized, collaborative, and nonproprietary; based on sharing resources and outputs among widely distributed, loosely connected individuals who cooperate with each other

[121] http://www.clickworkers.arc.nasa.gov/ [data of access: 05.11.2009].
[122] http://www.pgdp.net/c/ [data of access: 05.11.2009].

without relying on either market signals or managerial commands" (Benkler 2006: 60). He argues that its emergence is due to a change from the *industrial information economy* of mass media to the *networked information economy,* characterized by "[...] decentralized individual action—specifically, new and important cooperative and coordinate action carried out through radically distributed, nonmarket mechanisms that do not depend on proprietary strategies" (Benkler 2006: 3).

This third mode is not meant to replace markets and firms as the two main contemporary modes of production. It is rather a third type that emerges when certain *prerequisites* are met, i.e. ubiquity of networked personal computers, and it is particularly *advantageous for certain types of products,* namely intellectual products, such as information, culture and knowledge. This means that while commons-based peer production might be better suited to create software or to write an online encyclopedia, it might not be the best way to produce cars. However, since the focus of this thesis lies exactly on one of those intellectual products, namely knowledge, analyzing the characteristics and prerequisites of CBPP may be instructive for the development of a new socio-epistemological framework to analyze epistemic social software.

Organizing Collective Action: Allocation, Division, Integration & Motivation

Similar to markets and firms, in CBPP the problem of how to organize collective action has to be solved. While firms and market operate on managerial commands, contracts and property rights, CBPP solves this problem differently. Take a classic example of CBPP: the open source movement. In this community thousands of programmers collaborate voluntarily to produce software that they and others can use freely. That such a loose network of programmers can compete with the R&D departments of the biggest international software companies should be reason enough to analyze these collaborations from a socio-epistemological point of view. To understand why and how CBPP functions, basically three types of question have to be asked. First, how is collective action organized? How are people allocated to tasks? Second, how are tasks split up and how can they be (re-) integrated again? Thirdly, how are people motivated to take part in these collective endeavors and how could such motivation be sustained. These questions should sound familiar, and indeed, similar issues have been addressed in socio-epistemological considerations on the distribution of epistemic labor as outlined in the previous chapters. Given the different theoretical background of Benkler's analysis, i.e. mainly transaction cost analysis, a comparison of the answers given and the conclusions drawn should be rewarding.

Allocation of Epistemic Tasks

In order to organize large scale collaboration without market signals or managerial commands, CBPP basically depends on *self-attribution with checks and balances*. In contrast to firms, where tasks are often assigned to others, CBPP is based on self-assignment of tasks: programmers in open source projects or Wikipedia authors choose for themselves whether, how and to what extent they want to participate in their project.

On the one hand, such self-attribution can solve allocation problems quickly, because instead of managers searching for the right person for a given task, people will simply indicate themselves whether they are able and willing to do this task. Benkler argues that, from the perspective of transaction costs, "[t]he widely distributed model of information production will better identify who is the best person to produce a specific component of a project, all abilities and availability to work on the specific module within a specific time frame considered" (Benkler 2002: 414).

However, people are not always the best judges of their abilities. As John Hardwig has stressed, the necessity of adequate epistemic self-assessment for scientific collaboration is fundamental (Hardwig 1991). The same surely holds true for epistemic collaborations more generally. Given that this problem exists in science as well, it seems a bit precipitous to give up on epistemic self-assignment only because some people might misjudge their competencies.[123] Just because one cannot be 100% sure about the competency, honesty and adequate epistemic self-assessment of one's peers does not imply that one could or should refrain from collaborating. Rather, control mechanisms are needed to detect whether someone lies, is incompetent and does not even know he is incompetent. In science, *peer review* emerged as the most important quality control mechanism.[124] In CBPP, peer review is also one of the crucial control mechanisms, but it is only one among others. Depending on the project there is a variety of different

[123] Another problem is that with self-attribution it is likely that not all tasks will be done. If no one is willing to do something voluntarily, a task may remain unfulfilled. I return to this issue in Chapter 11 when comparing academic peer review and Amazon rating systems.
[124] For an analysis of the history, the pros and cons of benefits as well as possible alternatives, please confer for instance Wakeling et al. 2010.

mechanism employed to weed out low quality. On platforms such as Slashdot.org and Kuro5hin processes similar to traditional peer review are employed. In the case of NASA clickworkers, quality is ensured through statistical weeding out and the use of redundancy for verification.

The reference to statistical weeding out and the potential benefits of redundancy leads to the question of size. According to Benkler, when it comes to CBPP bigger is better - up to certain boundaries.[125] He states that "[p]eer production relies on making an unbounded set of resources available to an unbounded set of agents, who can apply themselves toward an unbounded set of projects. The variability in talent and other idiosyncratic characteristics of individuals suggests that any given resource will be more or less productively used by any given individual and *that the overall productivity of a set of agents and a set of resources will increase when the size of the sets increases toward completely unbounded availability of all agents to all resources for all projects*" (Benkler 2002: 415f, emphasis added).

Dividing and Integrating Epistemic Tasks

One crucial prerequisite for the success of collaborative endeavors concerns the structure of the tasks. Clearly, in order to collaborate on something, tasks have to be distributable over different people and therefore have to be divisible in the first place. Some tasks are more easily divisible than others. An encyclopedia can be more easily distributed over many people than a text book. This is the case not only because an encyclopedia is already structured into different articles and the different roles are more clearly separable, encyclopedias also demand a lesser degree of coherence than high school textbooks (Benkler 2006: 326). One might think that the more complex a project is, the more difficult its division and hence also collaboration. However, Benkler argues that "[p]eer production is limited not by the total cost or complexity of a project, but by its modularity, granularity, and the cost of integration" (Benkler 2002: 435).

Modularity refers to the sheer fact of whether and to what extent tasks are divisible into subtasks. Some tasks can be split up more easily than others and accordingly some tasks are better suited for collaboration than others. *Granularity* refers to the different sizes of

[125] The extent to which bigger is better depends on the rise of transactions costs, the danger of useless duplication of effort (as opposed to useful redundancy) and degree of standardized effort (versus variability of talent). (Benkler 2002: 422)

modules. In open source projects and in Wikipedia people can chose the extent to which they are willing to participate. They can choose to do more or less, depending on the time they are willing to devote. If people can choose different types and sizes of modules, it is much more likely that they will continue to participate a bit even in stressful times and to participate more in times of boredom.

Once tasks have been split up, they need to be combined again in the end. Thus, a crucial aspect concerns the *costs of integration*. The larger and more modular a project is, the higher are the costs of integration. Therefore, the ability of CBB to efficiently and successfully integrate individual contributions into a whole is one crucial factor often deciding upon its success or failure. Accordingly, Benkler argues that "[...] for a project to be susceptible to sustainable peer production, the integration function must be either low-cost or itself sufficiently modular to be peer-produced in an iterative process." (Benkler 2002: 436). Different social and technical solutions to the problem of task integration exist and are used in different projects. He distinguishes four strategies for information integration in CBPP: iterative peer production of the integration function itself, technical solutions embedded in the collaboration platform, norm-based social organization, limited reintroduction of hierarchy or market to provide the integration function alone (Benkler 2002: 436). Different projects employ different mechanism or combinations thereof; Wikipedia, for instance, combines all four types.

Motivation of Participants

The last crucial question is how to get people motivated to participate and how to keep them motivated over time. While in firms and markets, financial profit clearly is one major motivator; financial incentives are usually not decisive in CBPP. Skeptics of CBPP sometimes argue that projects based solely on altruism are doomed to fail on the long run. However, such an appeal to altruism might not even be needed. It is rather that motivations are more diverse than just bound to financial gain. Take academia. I do not attempt to paint too rosy a picture of academia or even downplay the crucial role of basic income. Nonetheless, many activities that academics undertake cannot be explained by simple reference to monetary reward. Academics for the most part do not get extra payments for publishing or presenting their work, for reviewing, for supervising theses. All these tasks are compensated for in their salaries, provided they have an appointment. However, this does not mean that scientists are pure altruists or unflinching chasers of truth either. Scientists may be motivated by the reputation of their peers. They might also strive for truth. They may value the freedom of research. They might of course also be motivated by future financial gain through a better position (or an appointment at all to begin with). Obviously, there is a mix of different motivations for action in academia, and as Philip Kitcher has shown, such epistemically sullied agents are not detrimental to science (Kitcher 1993). Quite to the contrary, a diversity of differently motivated people might actually have positive effects on science by avoiding premature consensus. Taking part in collective voluntary endeavors on the Web is based on rather similar motives than taking part in science, although Benkler uses more positive terms than Kitcher to describe such diversely motivated agents. What motivates people, Benkler argues, is in all probability a combination of different types of motivation: the pleasure of creation, reputation gains and also much more mundane benefits such as the acquisition of skills that might help later on in finding a job – all these motivators and others may play a role in various combinations (Benkler 2006).

Besides this more general question of what motivates people to act voluntarily, there is also the specific question of how to design a project so that people are motivated to participate – and continue to be motivated over time. According to Benkler, in the light of motivational issues, granularity and modularity of tasks may play a major role for the *initial motivation* to participate, while the effectiveness of integration rather *keeps people motivated*. If people cannot choose to what extent and in what way they want to participate or get the impression that their work is wasted because of inefficient or lacking integration, their motivation to participate will in all likelihood drop. Moreover, CBPP communities are often quite sensitive to misuse, such as unilateral appropriation or economic exploitation more generally. If Wikipedia decided to sell their information, this would probably be its death sentence. In a similar vein, the likelihood that people will donate processing power or memory capacity of their home computers for distributed computing depends crucially on the perceived value of the project. In a presentation on distributed systems, Adam L. Beberg from the Stanford Computer Science Department, summarized the first rule of distributed computing with a quite innocent request: Don't be evil![126] People participate in voluntary, collective action only

[126] The talk was entitled „Distributed Systems: Computation with a million friends (and a few foes). It was delivered on April 30th. 2008 at the Stanford EE Computer Systems Colloquium. (http://www.stanford.edu/class/ee380/Abstracts/080430.html)

if the goals of the project are transparent and good for humanity according to their views.

Benkler: Conclusions & Socio-Epistemological Relevance

Which conclusions can be drawn from Benkler's notion of CBPP for a socio-epistemological analysis of epistemic social software and why is it relevant for a social epistemology for socio-technical epistemic systems? First of all, Benkler not only analyses similar systems as the one's addressed in this thesis, he also asks similar questions and raises similar topics. Since Benkler's analyses are rooted in a very different disciplinary background, comparing his insights and results with socio-epistemological ones may thus serve as a probing stone for the validity of a new socio-epistemological framework.

Moreover, the relevance of CBPP for social epistemology is even strengthened by the *similarities between science and CBPP*, which Benkler notes himself. Indeed, science serves for him as a prime example for commons-based peer production. He argues that science is a classic case of commons-based *peer production* because in science "[t]housands of individuals make contributions to a body of knowledge, set up internal systems of quality control, and produce the core of our information and knowledge environment." (Benkler 2002: 382). And it is *commons*-based peer production because its product – knowledge and information – usually is a commons that has to be made publicly available.

Although crucial differences between science and certain examples of CBPP have been assessed (e.g. the differences between Wikipedia and science outlined by Wray 2009), science and CBPP nonetheless share many features: both are usually *non-proprietary* and *non-market modes of production*. Moreover, both entail a *division of cognitive labor* and as such depend on the competency, honesty and adequate epistemic self-assessment of its participants (cf. Hardwig 1991). Both have developed *mechanisms to control* these factors, such as peer review, although the control mechanisms employed in different projects might differ. Finally their contributors are motivated by similar *motives*: pleasure of creation, reputation gains and more mundane gains.

Benkler's view on the *role of technology* in human affairs is also similar to the one adopted in this thesis. Benkler acknowledges the centrality of technology for the societal change without falling victim to technological determinism. With reference to Barry Wellmann's usage of the term "affordances" and Langdon Winners "political properties of technologies" Benkler (Benkler 2006: 17) aligns himself with certain views concerning the relationship between the technical and the social from the field of Science and Technology Studies portrayed in Chapter 3. He states that "[d]ifferent technologies make different kinds of human action and interaction easier or harder to perform [...]" and concludes that "[...] neither deterministic nor wholly malleable, technology sets some parameters of individual and social action" (Benkler 2006: 17). With distinct reference to the *triplicate relationship between the social, the technical and the epistemic,* he asserts that "[t]he actual practices of human interaction with information, knowledge, and culture and with production and consumption are the consequence of a feedback effect between social practices, economic organization, technological affordances, and formal constraints on behavior through law and similar institutional forms" (Benkler 2006: 17). I fully agree with Benkler on his view on technology. Yet given that we analyze the same phenomena, this accordance is probably not too surprising.

Finally, Benkler's approach goes beyond a mere description of information production in economic terms. Instead he provides a *moral framework* to understand the profound changes brought about by the availability of networked computers and he takes a *normative stance*. In addition to this quite more general moral or normative thrust, Benkler delivers crucial insights and *criteria* to evaluate different socio-technical systems. He shows what is necessary for commons-based peer production to function well. The criteria for such functioning center around issues that have been outlined before: the different epistemic functions of communication for peer production (uttering content, relevance/accreditation and distribution) and the organization of collective action (allocation of tasks to agents, division and integration of tasks, motivation) (Benkler 2006: 68ff). These criteria can be used not only for the *evaluation of existing projects*, but also for the *design and developments of new socio-technical systems for epistemic purposes*. To recapitulate, to promote commons-based peer production systems should meet the following prerequisites:

1. The tasks should be highly modular and of varying granularity. With rising scale, the relevance of both modularity and granularity increases.

2. The integration of tasks should be low-cost, possibly partly automated and/or again modularized
3. Self-attribution of tasks should be combined with robust mechanisms for quality control. There are various social and technical mechanisms that can be used and also combined for quality control.
4. A diversity of motivations for participation is possible and possibly beneficial in its diversity.
5. To attract a vast crowd the goals of the project should be noble.
6. To keep the crowd on board, economic exploitation, unilateral appropriation has to be avoided at all costs and transparency has to be ensured.

I think these criteria are worth remembering for the development of a socio-epistemological framework to analyze epistemic social software and socio-technical epistemic systems more generally.

8.2.2 Surowiecki's Wisdom of the Crowds

In his book "The wisdom of the crowds - why the many are smarter than the few" (Surowiecki 2004), Surowiecki analyzes collective ways of finding solutions to problems or making predictions. The term *wisdom of the crowds* has ever since become a synonym for the success of collective epistemic processes on the Web. Surowiecki's major argument is that "under the right circumstances, groups are remarkably intelligent, and are often smarter than the smartest people in them" (Surowiecki 2004: XIII). As a consequences, it is epistemically superior to ask the crowds instead of wasting time to chase the expert (Surowiecki 2004: xv).

In the context of Web2.0, the term *wisdom of the crowd* and Surowiecki's argument are frequently used to argue for the epistemic merits of bottom-up, democratic approaches as compared to more meritocratic systems depending on the differentiation between lay people and experts. This anti-expert-rhetoric on the Web has stirred a lot of debate and in this debate the impression is conveyed that one has to choose between scale and expertise as two distinct options for problem solution: either you ask a single expert, who knows it all or you use statistical averaging of people who do not know nothing at all. While more nuanced analyses of the respective strengths and weaknesses of different forms of problem solution seem to be lacking, simplifications and misrepresentation often lead to exaggerated claims. While some authors proclaim the end of experts as we know them (Ayres 2007), others send out warnings about "The Cult of the Amateur" (Keen 2008). The truth may lie somewhere in between as Surowiecki notes himself when writing that "[...] for the group to be smart, there has to be at least some information in the 'information' part of the 'information minus error' equation" (Surowiecki 2004: 10).

It is however not only rhetoric and the reception of a term in the Web2.0 discourse, which makes Surowiecki's work central for this thesis. Although Surowiecki describes different modes of problem solving as well as different kinds of problems that can be solved collectively, the term "wisdom of the crowds" is primarily used to describe *one very specific mechanism* which consists in a combination of *individual betting, aggregation and averaging* for decision making and prediction. And the Web2.0 is full of socio-technical systems that function exactly in such an aggregational mode. Recommender Systems, websites like Digg.it, Google's Page Rank algorithm – in all of these examples and many more – masses of individual votes are being aggregated for epistemic purposes. Indeed, aggregation is one of the major epistemic closure mechanisms employed on the Web. I argue in detail in Chapter 10, it forms one of three generic socio-epistemic closure mechanisms that should be distinguished to analyze socio-technical epistemic systems.

Surowiecki's book is divided into two parts. In the first one Surowiecki explicates his theoretical assumptions (Surowiecki 2004). More specifically, he argues that for the wisdom of the crowds to come to the fore, three prerequisites have to be met: groups have to be *diverse*, their members have to act *independently* and *decentralization* has to be *coupled with adequate aggregation mechanism*. I explicate these criteria below in some detail, because they are crucial for formulating prerequisites for one type of epistemic sociality I distinguish in my model: ES:Aggregation.

After devoting a chapter to each of these prerequisites Surowiecki differentiates *three different types of problems* that can be collectively solved: *cognition problems, coordination problems* and *cooperation problems.* For my own socio-epistemological framework, Surowiecki's analyses on *cognition problems* are central. For this type of problems, he argues that a combination of *individual betting, aggregation and averaging* is superior to other epistemic strategies when it comes to decision making and prediction (Surowiecki 2004). In the second part of the book Surowiecki shows how the wisdom of the crowds functions in a variety of different fields by providing brief anecdotes on topics such as traffic regulation and jury formations, as well as different processes within companies, markets, democracies and science.

Aggregating to Know

In the introduction to the book, Surowiecki uses two main examples to elucidate the wisdom of the crowds (Surowiecki 2004: xiff): how it helped Francis Galton to judge the weight of an ox and how it helped the US navy to find a lost submarine.

Francis Galton, one of the founders of psychometrics did the following real-life experiment. When visiting a country fair in 1906, he observed a popular game of those days, the weight-judging competition. After buying a ticket, the visitors of this fair could place a bet on the weight of an ox and whoever came closest to the ox's weight won a price. About 800 people submitted their guess and they were a pretty diverse crowd. There were butcher, maids, farmers, children, etc. One would expect that the guesses of the butchers and farmers might be best, because they should experts on the weight of ox, while maybe the village's priest, the clerk or the maid would know less about the weight range of cattle. Galton was interested in the abilities of the *average person*. This interest for him was also of political nature, because he wanted to assess the capabilities of average voters in making plausible judgments. To assess the average person, Galton collected the bets and calculated the average of all the weight guesses. This average of all judgments, he thought, would be the *average person's bet.* Clearly, since Galton was not too optimistic about the capabilities of the average man, especially given that there were so many non-experts in the crowd of betters, he assumed that this average bet would be far off. He was surprised. The averaged guess of the crowd was 1,197 pounds – and the real weight of the ox was 1,198 pounds. Surowiecki cites Galton stating that "[t]he result seems more creditable to the trustworthiness of a democratic judgment than might have been expected" (Surowiecki 2004: xiii). And Surowiecki adds: "That was, to say the least, an understatement" (Surowiecki 2004: xiii).

In the second example, Surowiecki describes a naval officer in search of a lost submarine who came up with the idea to have people *bet on different scenarios* of what had happened to the submarine instead of having people *discuss* the issue. After having developed these scenarios, the officer aggregated these bets by using Bayes' Theorem. By aggregating these bets according the this algorithm, he obtained a *collective* *judgment* about the location of the submarine – and this judgment was pretty close to where the submarine actually sank and enabled its location.

Judging the weight of an ox and locating a lost submarine are two examples of a problem type that Surowiecki labels *cognition problems* (Surowiecki 2004: xvii). In both examples, the aggregation of individual votes enabled predictions that were closer to the true value than individual bets. One commonality of these two examples is that there are *true values* to which the predictions could be compared: the ox was weighted, the submarine was found. Hence, definite solutions to the problems existed and the predictions could be tested and verified. Cognitive problems, however, also include more open-ended question, such as finding the best place to build a new public swimming pool as well as probabilistic questions, such as the likelihood of the approval of a drug by the Federal Drug Administration (Surowiecki 2004: xvii).

Surowiecki distinguishes such cognition problems from *coordination and cooperation problems*. According to Surowiecki, *coordination problems* "[...] require members of a group (market, subway riders, college students looking for a party) to figure out how to coordinate their behavior with each other, knowing that everyone else is trying to do the same" (Surowiecki 2004: xviii). Cooperation problems, by contrast, for Surowiecki, "[...] involve the challenge of getting self-interested, distrustful people to work together, even when narrow self-interest would seem to dictate that no individual should take part" (Surowiecki 2004: xviii). Surowiecki delivers an abundance of examples for different – more or less successful - solutions to various coordination and cooperation problems. Although his examples come from a variety of different fields and his perspective is different from those authors portrayed in the previous chapters, his comments on the epistemic relevance of coordination and collaboration correspond quite well to socio-epistemological considerations, as becomes obvious in his chapter on science.

Reminiscent to Kitcher's (Kitcher 1993) and Solomon's (Solomon 2001) works, Surowiecki adopts a similar macroscopic perspective and analyzes the ways in which science as a collective endeavor is organized. One of the examples he chooses is the distributed research effort that revealed the causes of SARS. Surowiecki uses this large-scale international collaboration to show how a combination of competition and collaboration conducted by independent researchers, who exchange information, lead to incredibly fast research results. Moreover, as Thagard (Thagard 1997a), Hardwig (Hardwig 1985; Hardwig 1991) and others, Surowiecki argues that modern *science is collaborative* in nature and stresses the role of *trust* for science (cf. also Welbourne 1986, Shapin 1994 and Kusch 2002). Similar to Yochai Benkler's comments on information and *knowledge as public and non-rival goods* (Benkler 2006: 36f), Surowiecki stresses the distinctiveness of knowledge and information as goods, whose value does not diminish through transmission and emphasizes that *free access to information* is crucial for science.[127] He argues that although science can flourish in private companies, the freedom of science and knowledge might be endangered by rising *commercialization*, a point that has also been made – and discussed - by various philosophers of science (e.g. Ziman 2003). Finally, while he acknowledges the crucial role of *reputation* as a motivator and currency in science, he also stresses the dangers that may come with an overreliance on reputation as a proxy for assessing the quality of content by referring to the *Matthew-Effect* (Merton 1988).

Although based in very different theoretical frameworks, the insights Surowiecki obtains from his analyses on coordination and cooperation problems correspond well with many claims and insights from social epistemology. Hence, one can conclude that addressing these issues is important not only to analyze science, but socio-epistemic practices and systems more generally. Yet while his insights on *coordination* and *cooperation problems* may merely support socio-epistemological claims made before, his analyses of *cognition problems* are far more innovative and relevant for this thesis. Indeed, the mechanisms of individual betting and statistical aggregation of results form the basis of many socio-technical epistemic systems. Through such aggregational mechanisms, new modes of knowledge creation have emerged on the Web. Since this thesis aims at providing a framework to analyze such systems and the processes taking place within them, getting a better understanding of the functionalities, the prerequisites, the benefits and shortcomings of aggregational mechanisms is essential. Surowiecki's work on cognition problems deliver a good starting point for the development of a framework to analyze aggregation as a socio-epistemic mechanism of closure to be addressed in Chapter 10.

[127] Please note though that Surowiecki (2004: 168) still argues that patenting and copyright may be a necessary motivator for creativity, while Benkler (2006) clearly refutes that claim.

The Core of the Wisdom of the Crowd: Aggregating Individual Judgments

What is the core of the wisdom of the crowds? Think back to the examples depicted before: Galton's ox and the sunken submarine. In both cases – and in all other examples Surowiecki uses - *individuals independently made guesses*, which were then *aggregated* statistically. This aggregated result was then taken to be the collective judgment of the group. In the Galton example, the aggregation process was quite simple and consisted in simple averaging: add up the guesses, divide by the number of betters. In the case of finding the sunken submarine, the aggregation mechanism was more advanced and included a weighted aggregation based on Bayes' Theorem. One application of this type of epistemic sociality that has attracted a lot of attention recently – decision or prediction markets – are yet another refinement of the same process of individual guesses combined with smart stochastical aggregation mechanism.

Surowiecki shows that in many cases, decision making, predictions and problem solutions based on stochastical aggregation of individual bets outperform individual judgments even if they are made by experts. Moreover, under certain circumstances, such a mode of aggregation often leads to better results than deliberative processes. In other words in certain situations it is better to let people bet individually on an issue and calculate their average bet than letting them discuss the same issue. However, for *group intelligence* or *wisdom of the crowd* based on such a mechanism to occur, certain prerequisites have to be met: the groups have to be diverse, people have to be independent and there have to be adequate aggregation mechanisms.

The rationale behind these claims for a superiority of aggregated judgments made by a diverse crowd of independent individuals is a purely statistical and related to the axioms of classical test theory. Simply put, the idea is that if one repeats a measurement the errors will cancel each other out and one will be left with the information only. As Surowiecki states, at the heart of his argument lies a "mathematical truism. If you ask a large enough group of diverse, independent people to make a prediction or estimate a probability, and then average those estimates, the errors each of them makes in coming up with an answer will cancel themselves out. Each person's guess, you might say, has two components: Information and error. Subtract the error, and you're left with the information" (Surowiecki 2004: 10). However, to make sure that you really end up with the information – or the true value - this mechanism is based on various prerequisites. Not only do you need a *large enough number* of measurements. The measurements also need to be *independent* to avoid systematic bias.

Diversity

That cognitive diversity is beneficial for socio-epistemic processes has been argued extensively within social epistemology (e.g. Longino 2002c, Solomon 2006) and other fields of research (e.g. Page 2007). Explicitly referring to early works of Scott Page (Page 2001), Surowiecki argues that the more diverse a crowd is, the more likely it is that they come up with a diverse set of alternative problem solutions. However, diversity is not only needed to uncover possible alternatives, but also to choose among them. To use Benkler's (Benkler 2006) terminology, diversity is not only used to "utter content", but also for relevance assessment and accreditation. For the assessment of different solutions diversity matters, because it offers a variety of different perspectives and helps to avoid *groupthink*, a cognitive failure in homogenous groups, which "works not so much by censoring dissent as by making dissent seem somehow improbable" (Surowiecki 2004: 37).

Diversity is related to *scale*, since the bigger the group is, the more likely is cognitive diversity. In smaller groups, however, diversity might have to be induced artificially, e.g. by certain employment policies or knowledge management measures. Surowiecki links his request for diversity with a critique of *expertise* in three ways. First of all, he argues that expertise usually refers only to a very narrow realm. By combining this specificity with the second argument, that experts tend to be overconfident, i.e. lack one of Hardwig's crucial prerequisites for successful collaboration, namely "adequate epistemic self-assessment"(Hardwig 1991), Surowiecki uses this as an argument why as a general practice, reliance on expertise is less advantageous than a wisdom of the crowds (Surowiecki 2004: 31ff). Finally, even *groups of experts* might be inferior to groups mixed of experts and non-experts, because they are less diverse. This argument is based upon the idea that experts use similar cognitive strategies and might thus not come up with new, alternative ideas if their usual strategies are not fruitful for a given task.

Independence

Diversity is crucially related to independence, because diversity can help preserve independence (Surowiecki 2004: 39). From an epistemic point of view, independence is needed to avoid two major cognitive failures: *group think* and *information cascades*. *Group think* describes a situation in which the generation of alternative problem solutions is unlikely because of a lack of cognitive diversity and deliberation leads to additional conformity pressure (Surowiecki 2004: 337ff). As a consequence Surowiecki advocates for voting and betting procedures over deliberative means to reach consensus.

Information cascades are a different problem and they result from decisions taken in sequences rather than simultaneously. According to Bikhchandani, Hirshleifer et al. (Bikhchandani, Hirshleifer et al. 1992: 992) "[a]n informational cascade occurs when it is optimal for an individual, having observed the actions of those ahead of him, to follow the behavior of the proceeding individuals without regard to his own information". An example would be a situation that might sound familiar. Imagine you walk around in Paris in search for a good restaurant and you do neither have a guide book nor any other information on the respective quality of restaurants you pass by. A useful heuristic may be to use the number of guests as a proxy for the quality of a restaurant. While inferring the quality of the food from the popularity of a restaurant indicated by the number of guests might work out quite well in practice, it depends on the fact that some of those who go there actually know that the food there is good. Imagine that there are two restaurants next to each other. Let's assume both look nice and have the same menus. Yet while one of them offers fabulous food, the food in the other restaurant is always salty and the waiters unfriendly. For the sake of simplicity let the good restaurant be named "Stairway to Heaven" and the bad one "Highway to Hell". When the first guests arrive and stop in front of the two restaurants, they cannot decide based on the heuristic or number of guest = quality. And without any knowledge about the restaurants, they simply have to pick one of the two randomly. If they end up in the "Highway to Hell", this may be just bad luck. However, if every person decides upon the number of guests this night this will lead to everybody having tasteless food at the "Highway to Hell", while the "Stairway to Heaven" is empty. Clearly, something went wrong here. This simple example offers an example of a full-fledged information cascades with clearly negative results for all parties involved except from the owners of the "Highway to Hell". To use Surowiecki's terms "everyone ends up making the wrong decision, simply because the initial diners, by chance, got the wrong information"

(Surowiecki 2004: 53). Instead of aggregating information, basically no information is used in such cascades: the initial decision was based on an error and everyone repeats the same mistake, by assuming that those before *knew* – which they did not. Where private information should be aggregated, ignorance is simply spread and reinforced. To avoid such information cascades diversity and independence of judgment has to be ensured. One of the easiest ways – at least in laboratory situations – is to ask people to take their individual decisions simultaneously instead of sequentially.[128]

Decentralization & Modes of Aggregation

The benefits of decentralization are implied in Surowiecki's request for independent judgments of diverse agents. However, if these independent tasks or judgments do not get integrated properly again, information gets lost and the intended benefits of decentralization may be lost. Surowiecki summarizes the dilemma as follows: "Decentralization's great strength is that it encourages independence and specialization on the one band while still allowing people to coordinate their activities and solve difficult problems on the other. Decentralization's great weakness is that there's no guarantee that valuable information which is uncovered in one part of the system will find its way through the rest of the system" (Surowiecki 2004: 71). Thus, effective modes of integration are needed. There are of course different, non-aggregational methods of integrating individual epistemic results into collective epistemic results, such as the ones outlined in the next chapter. Surowiecki's focus however, clearly lies in such aggregational mechanisms.

The simplest method of information aggregation proposed was simple averaging. While this process may be sufficient for tasks as simple as judging the number of jelly beans in a jar or the weight of Galton's ox, more complex problems may require more complex aggregation mechanisms. The case of the sunken submarine would be one example where different aspects and criteria have to be weighted and hence Bayes' Theorem was chosen.

Various more complex aggregation mechanism are also employed in *decision* and *prediction markets* (Surowiecki 2004: 79ff). In an article that surveys and evaluates the merits and limits of such markets, Bragues defines prediction markets as "venues in

[128] For a critique of such simple recommendations to avoid information cascades please confer Coady (2006).

which individuals trade securities whose value is tied to the outcome of a future event" (Bragues 2009). Some examples of such decision markets that Surowiecki refers to are Iowa Electronic Markets (IEM), in which election results are predicted and FuturePAM, resp. the Policy Analysis Market PAM, where bets were made on international affairs (Surowiecki 2004: 79ff). The fact that people could bet on terrorist attacks or assassinations has stirred a lot of moral debate, so that the project was terminated. The epistemically interesting question is, however, whether and why such markets function better than traditional polls or expert predictions in forecasting the future. It has been argued that "prediction markets are the opposite of social collaboration" (Bragues 2009: 95). Participants *do not collaborate*, but *compete* against one another. Each person independently makes his or her bets and competes against the others in their effort to make money. Although people compete against each other, it is their *combined* effort, an *aggregated* result that is the epistemic outcome of a prediction market. Thus, decision markets are an example of a distinct type of epistemic sociality in which the results of individual independent action get aggregated to reach collective epistemic results.

Surowiecki: Conclusions & Socio-Epistemological Relevance

What insights can be taken from Surowiecki's analyses for the development of a socio-epistemological framework to analyze epistemic social software? The main reason why I have inserted this section on Surowiecki's is that his exploration in the wisdom of the crowds delivers crucial insights for one of my three types of epistemic sociality*: aggregation*. Since I consider aggregation, together with *integration* and *selection*, to be a generic mechanism for closing socio-epistemic processes, Surowiecki's analyses concerning the prerequisites for the proper functioning of aggregational mechanisms can serve as criteria for the analysis and evaluation of socio-technical epistemic systems. A proper socio-epistemological analysis of systems employing aggregational mechanisms has to check whether independence and diversity of participants is ensured and whether adequate aggregational algorithms have been employed. Since aggregational systems have been on the rise on the Web in the last years, being able to specify criteria to analyze such systems is an important aspect of any socio-epistemological framework to analyze such systems.

Besides adopting Surowiecki's criteria, another important issue can be *inferred* from his observation that the emergence of any wisdom of the crowds crucially depends on the aggregation of individual results. Without aggregation, no epistemic results are obtained. The creation of knowledge qua aggregation is delegated to algorithms. Given the centrality of these aggregation processes it becomes obvious that a lot of trust needs to be placed in such aggregational mechanisms, i.e. authority is attributed to algorithms in this type of epistemic sociality. In aggregational socio-technical epistemic systems trust is deferred from individual epistemic agents to the mechanism of closure, from the people to the algorithm, from human to non-human epistemic agents. This shift in trust and authority is a central feature of aggregational systems, as I argue in more detail in Chapter 10. But trust can be misplaced; authority can be attributed inappropriately. As a consequence, this new locus of trust and authority has to be made subject to scrutiny by any social epistemology to assess the validity of this delegation of epistemic trust and authority.

8.3 The Core: Closure through Integration, Aggregation or Selection

In the following chapters I outline my socio-epistemological framework to analyze epistemic social software understood as socio-technical epistemic systems consisting of multiple agents. This framework is based on a *tripartite classification of the different socio-epistemic mechanisms of closure* employed in such systems, i.e. I argue that there are *three generic mechanisms of closing socio-epistemic processes* in which multiple agents are involved. These mechanisms are *integration, aggregation* and *selection*. By distinguishing systems according to the mechanisms of closure they employ, important differences between systems come to the fore. Hence, I use the names of the closure mechanisms to denote three types of epistemic sociality: Epistemic Sociality^I (ES^I), I for *integration*, Epistemic Sociality^A (ES^A), A for *aggregation,* Epistemic Sociality^S (ES^S), S for *selection.* These three types of epistemic sociality form the basis of the new socio-epistemological framework I propose to analyze epistemic social software and socio-technical epistemic systems more generally.

The term e*pistemic sociality* refers to the fact that in all cases there is *more than just one epistemic agent involved in the epistemic process*. Different concepts have been introduced to label and describe the utilization of a multitude of epistemic agents for epistemic tasks: *collective intelligence, wisdom of the crowds, team work, cooperation, collaboration*, to name just a few examples. I use the term *epistemic sociality* as a generic term that encompasses all these notions. The reason for using a generic term here is that each of the before mentioned terms, collective intelligence, wisdom of the crowds, etc., already implies a *specific* way in which the multitude of epistemic agents is organized. Yet those specificities are neither always clear, nor consistent, nor used in the same way by different authors. Hence, the terminology (ES^I, ES^A, ES^S) is introduced to avoid confusion. Nonetheless, reference is made to the common language terms introduced above to explain and describe the processes taking place in the three types of epistemic sociality before closure occurs.

Thus the utility of my analysis is twofold. First, providing a schema to distinguish different modes of epistemic sociality is of theoretical interest for social epistemology by offering a more fine-grained and nuanced framework for analyzing socio-epistemic practices. Differentiating distinct forms of epistemic sociality, outlining their prerequisites as well as their strengths and weakness is crucial for *understanding* the ways in which knowledge is or can be created socially. Secondly, such a classification can also serve as the theoretical basis for *improving* socio-epistemic practices and systems. Knowledge about the prerequisites, strengths and weaknesses of different social mechanism enables the design and development of socio-technical systems that serve epistemic purposes in a rational, efficient and effective way.

In particular, I can show that employing different mechanisms of closure has differential consequences for the organization of epistemic labour in socio-technical epistemic systems and for the relationships between different epistemic agents in such systems. With respect to these inter-agent relationships, I focus in particular on the differential placements of *trust* within such systems as well as on differential attributions of *authority*, among human epistemic agents and between human and non-human epistemic agents. Further, I highlight their respective strengths and weaknesses, as well as the epistemological and ethical opportunities and challenges they entail, drawing conclusions about how to best make use of each mechanism to design socio-technical epistemic systems that are epistemologically and ethically sound. Finally, I show that such mechanisms can be – and indeed are - combined in different socio-technical epistemic systems on the Web and elsewhere. Normative recommendations concerning systems design therefore do not consist in advocating one mechanism or another, but rather in finding useful combinations of socio-epistemic mechanisms of closure that minimize epistemic and ethical harm.

The following picture visualizes the basic structure of my model. Within socio-technical epistemic systems epistemic labor is distributed over multiple agents, which can either individually or collectively take part in epistemic endeavors. The epistemic results that they obtain depend upon the temporary closure of epistemic processes. Besides the termination of individual epistemic processes, i.e. the process of bringing content into a certain form an making in available to others, three generic social mechanisms of closure exist: integration, aggregation and selection. Systems employing these different mechanisms of closure differ with respect to numerous aspects. Most importantly, they differ with respect to the primary loci of trust, the attribution of authority, reputation and trustworthiness. They depend on different amounts and forms of independence and diversity. They are suited for different tasks and have different strengths and weaknesses. To outline these differences in detail is the goal of the next three chapters.

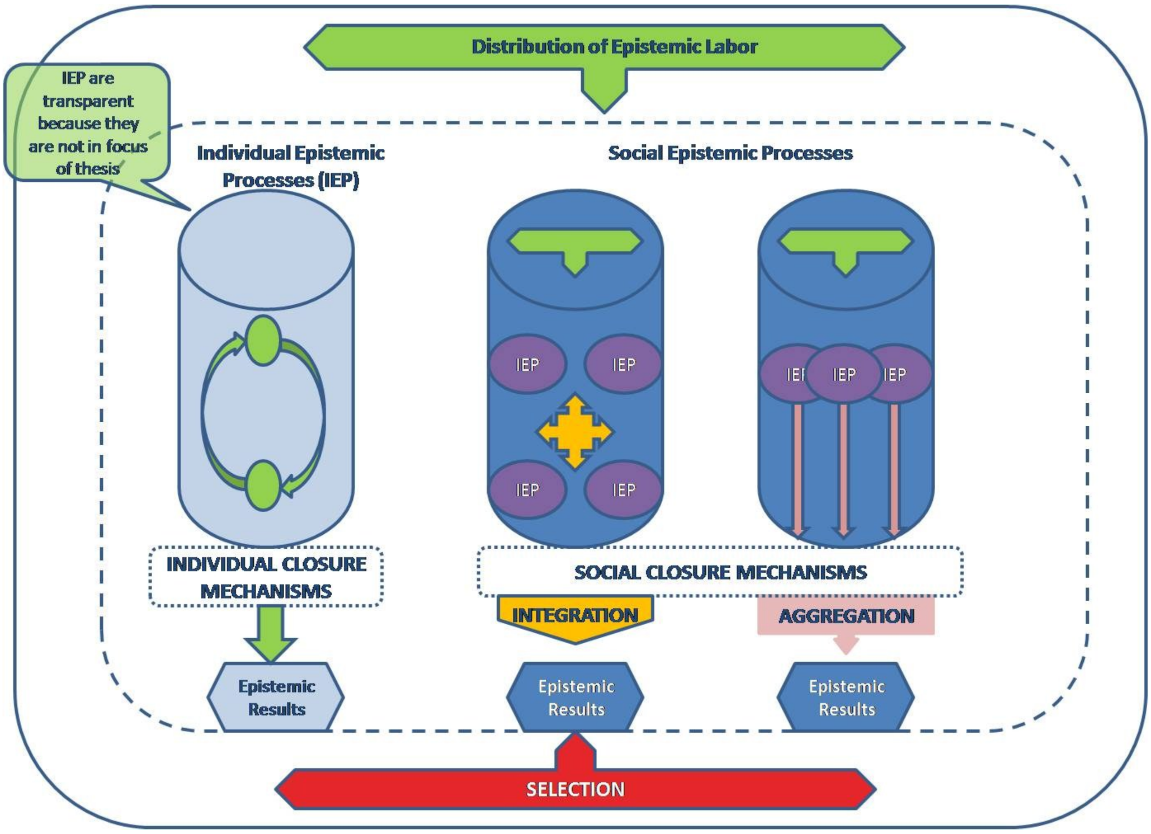

Figure 12: Types of Epistemic Sociality: ES:Integration, ES:Aggregation, ES:Selection

9 Epistemic Sociality 1: Closing Socio-Epistemic Processes Through Integration

9.1 Epistemic Sociality:Integration (ES^I)

ES:Integration (ES^I), is probably the most familiar form of epistemic sociality and refers to the type of epistemic practices that come to mind first when thinking about ways in which people collectively create knowledge. This type of epistemic sociality has also been at the heart of most works on the social aspects of science conducted in social epistemology as well as in other philosophical, sociological and historical analyses of science. Also on the Web, this type of epistemic sociality has also been studied most intensely, e.g. in Yochai Benkler's analyses of Wikipedia or the open source movement (Benkler 2006). In the following I portray Wikipedia as the prime example of this type of epistemic sociality on the Web in more detail and use LiquidPub, a research project in which I am involved as a case to exemplify ES^I in science.

9.1.1 ES:Integration & Wikipedia

Wikipedia is probably the single, most well-known example of ES:Integration on the Web. A multitude of people contribute and interact to create a collective epistemic product: a multi-lingual encyclopedia. Millions of epistemic tasks of varying kind and size are distributed over more than 85.000 diversely motivated volunteers via self-allocation. Yet how can we understand these epistemic processes? Since I have proposed a differentiation of types of epistemic sociality according to the mechanisms of closure, I first take a look at the integration of these innumerous epistemic tasks conducted by different agents. How is integration achieved in Wikipedia? In short, integration of epistemic content in Wikipedia happens on different levels with the help of *technical and social constraints*.

On a top level, Wikipedia as a whole is structured along languages. From the website of the Wikipedia organization http://wikipedia.org/ one can access different language versions, e.g. http://en.wikipedia.org/wiki/Main_Page, leading to the English version starting page, http://de.wikipedia.org/wiki/Wikipedia:Hauptseite, for the German starting page, etc. The differentiation between different languages is therefore indicated by the *en* for English, *de* for German, etc., followed by the common string

.wikipedia.org/wiki/. What comes after this common string indicates the topic of the article to which the URL is linking. For instance, http://en.wikipedia.org/wiki/Epistemology is linking to the article about epistemology or http://en.wikipedia.org/wiki/Macaron is leading to a description and depiction of a famous Parisian delicacy.

Hence, on the most global level the integration of different articles into the totality of the multi-lingual encyclopedia is first of all structured by this specific usage of URLs. There is a common top-level distinction into language versions, followed by a common string, followed by the indication and reference to the article. Only for the last part there is some variance: to create URLs indicating and linking to topics that need several words, people mostly use underscores (e.g. Main_Page), but some use colons (e.g. Wikipedia:Hauptseite), or combinations of colons and underscores (e.g. Portal:Featured_content).

From a socio-epistemic perspective it is essential that the integration of articles into the entirety of the encyclopedia depends on a combination of technical constraints and social constraints. *Technical constraints* are for instance: the necessity to create a URL, the necessity to indicate the language specifier, the necessity to use the common string, the necessity to complete the URL by creating an end that is not yet assigned and consists of valid characters.[129] However, with only those technical constraints there is still a huge variety of practices possible. Thus, the homogeneity of the URL-usage cannot be explained by technical constraints alone. For instance, it is technically feasible to use URLs consisting of very many characters, to use different characters to split words, or even to create URLs that do not content-wise refer to the topic of the linked article. However, given the uniformity of URL-design, obviously there are some rules, some conventions on how to create URLs. And indeed, there are guidelines for the creation of URLs to be found on Wikipedia. Besides proposals on how to link to article titles that consist of more than one term, numerous other recommendations are being

[129] Further information on technical constraints can be obtained here: http://en.wikipedia.org/wiki/Wikipedia:Naming_conventions_%28technical_restrictions%29 [date of access: 16.03.2010]

offered: on how to decide upon an article title in the first place, how to deal with common names, foreign names, how to be precise and unambiguous in labelling, etc.[130]

However, since most articles are written by more than one author, integration is not only crucial for integrating articles into the entirety of the encyclopedia, but also on a lower level, e.g. within each article. Content needs to be created by typing words into a template. Links need to be created, references should be made to literature, spelling and grammar needs to be checked. Many people conduct tasks of varying granularity and modularity, they spend different amounts of time in participation and they do this for different motivations. In the worst case, people may participate for the sole intention of destroying knowledge. Hence a system of checks and balances needs to be installed as well. This system of control is also based on social and technical features. Important technical features that serve to reduce vandalism are the history-page, which helps to track changes, and the possibility to easily undo revisions and restore content.

Yet in addition to the technical features, there are innumerous social constraints in Wikipedia. Indeed, a discourse analysis of all those norms, guidelines and policies may be a rewarding project. For instance, the Wikipedia Editing Tutorial[131] offers "guidance about the style and content of Wikipedia articles, and tell[s] you about the Wikipedia community and important Wikipedia policies and conventions". In addition to the Tutorial[132], which offers help on editing, formatting, linking, citing, etc, Wikipedia has a Manual of Style[133], and a Frequently Asked Questions section.[134] There are lists of *guidelines*[135] on behaviour, content, deletion, editing, naming, notability and style as well as lists of *policies*[136] on conduct, content, deletion, enforcement, legal, procedural and miscellaneous issues. After having completed the tutorial there is a link to numerous further sites subdivided into sections on "Advice and general information", "Policy references", "Editing references" and "Creating new articles".[137] I have provided a screenshot of this page just to give an idea about the breadth of social constraints that

---

[130] http://en.wikipedia.org/wiki/Wikipedia:Article_titles#Proposed_naming_conventions_and_guidelines [date of access: 16.3.2010]
[131] http://en.wikipedia.org/wiki/Wikipedia:Tutorial [date of access: 16.3.2010]
[132] http://en.wikipedia.org/wiki/Wikipedia:Tutorial [date of access: 16.3.2010]
[133] http://en.wikipedia.org/wiki/Wikipedia:Manual_of_Style [date of access: 16.3.2010]
[134] http://en.wikipedia.org/wiki/Wikipedia:FAQ [date of access: 16.3.2010]
[135] http://en.wikipedia.org/wiki/Wikipedia:List_of_guidelines [date of access: 16.3.2010]
[136] http://en.wikipedia.org/wiki/Wikipedia:List_of_policies [date of access: 16.3.2010]
[137] http://en.wikipedia.org/wiki/Wikipedia:Tutorial_%28Wrap-up_and_more_info%29 [date of access: 16.03.2010]

are communicated and fostered via these rules and guidelines. Obviously, social constraints, i.e. rules and norms, not only play a crucial role for the integration of individual contributions into the entirety of Wikipedia; these social norms which cover numerous pages are subject to discussions themselves. Moreover, the social norms were not decided upon once. Instead, they are performed and enacted through participation, they are fostered, strengthened or changed over time. They are created by what Kusch describes as *communal performatives*, the performative utterances carried out by a collective 'we' that take place in a distributed and fragmented manner (Kusch 2002: 17). They are what for science Longino describes as the *shared social norms* (Longino 2002c: 145). Obviously, such norms are not only crucial in science, but in any collective epistemic endeavour.

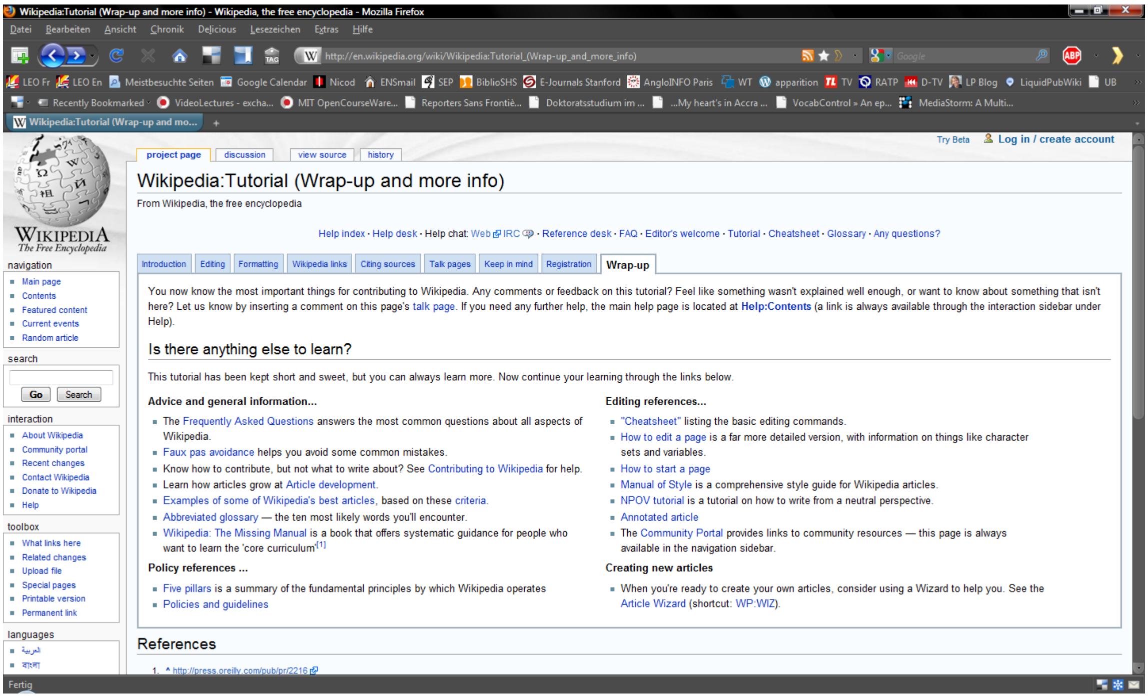

Figure 13: Social Constraints in Wikipedia[138]

9.1.2 ES:Integration & Academic Research

A field in which ES:Integration clearly also is the norm is *academic research*. Think about molecular biologists working in a research lab. Think about three pedagogical psychologists writing a textbook together. Think about an interdisciplinary team of scientists, ethicists, social scientists and legal scholars trying to outline the benefits and

[138] http://en.wikipedia.org/wiki/Wikipedia:Tutorial_%28Wrap-up_and_more_info%29 [date of access 16.03.2010]

problems around stem cell research together. Or think about a project funded by the European Union in which researchers with backgrounds in computer science and engineering, artificial intelligence, theoretical physics, economics, and philosophy try to come up with new ways of creating and scientific content. The last example refers to the project LiquidPub, a project funded by the Future and Emerging Technologies (FET) programme within the Seventh Framework Programme for Research of the European Commission (FET-Open grant number: 213360).[139] Since I am part of this project, I use it for some reflections upon the socio-epistemic processes taking place in interdisciplinary research projects in order to outline the functioning and some key elements of ES:Integration in science.

Given the focus of this chapter on collectively produced knowledge created via integration of individual contributions, I use the project description from the project's website instead of paraphrasing it.

> "The world of scientific publications has been largely oblivious to the advent of the Web and to advances in ICT. Even more surprisingly, this is the case even for research in the ICT area: ICT researchers have been able to exploit the Web to improve the (production) process in almost all areas, but not their own. We are producing scientific knowledge (and publications in particular) essentially following the very same approach we followed before the Web. Scientific knowledge dissemination is still based on the traditional notion of "paper" publication and on peer review as quality assessment method. The current approach encourages authors to write many (possibly incremental) papers to get more "tokens of credit", generating often unnecessary dissemination overhead for themselves and for the community of reviewers. Furthermore, it does not encourage or support reuse and evolution of publications: whenever a (possibly small) progress is made on a certain subject, a new paper is written, reviewed, and published, often after several months. The situation is analogous if not worse for textbooks.
>
> The LiquidPub project proposes a paradigm shift in the way scientific knowledge is created, disseminated, evaluated and maintained. This shift is enabled by the notion of Liquid Publications, which are evolutionary, collaborative, and composable scientific contributions. Many Liquid Publication concepts are based on a parallel between scientific knowledge artifacts and software artifacts, and hence on lessons learned in (agile, collaborative, open source) software development, as well as on lessons learned from Web 2.0 in terms of collaborative evaluation of knowledge artifacts."[140]

---

[139] Further information on the project can be found on this website: http://project.liquidpub.org/ [date of access: 16.03.2006]
[140] Taken from http://project.liquidpub.org/ [date of access: 16.03.2006]

As should be obvious from this description, the project itself addresses numerous socio-epistemic processes taking place within socio-technical epistemic systems. Moreover, this text and the website as a whole are examples of collectively produced epistemic results for which numerous epistemic agents have been responsible. When exploring the website, one comes across numerous epistemic products of varying size for which multiple researchers have interacted *with each other via technology* or *with technology* in various ways to create and evaluate content. However, these socio-technical epistemic processes, in which several agents are involved, need to be closed to lead to collective results, which can then be accessed on this website: products such as reports, demos, articles or links, which can be used as epistemic input for new socio-epistemic processes, in which researchers from the project and beyond can take part.

In the course of this three-year project more than 40 people with different competencies, skills and motivations have been contributing in varying degrees and various respects to the collective goals of the project by fulfilling epistemic tasks of different size and kind, by creating and co-creating epistemic products through closure of socio-epistemic processes. Within the project, partners interact and communicate not only with and about the new applications and technologies that are being developed. We also make extensive use of existing information and communication technologies to guide and support our interaction. We use various mailing-lists, GoogleDocs and GoogleCalendars, servers, a blog, a private and a public wiki, Delicious, CiteULike, to organize work, to interact and to communicate with each other.

Nonetheless, the goal of each of these communicative and interactive processes is the creation of tangible epistemic products: reports, articles and tools. For those reports and tools, the individual contributions have to be integrated. That is, the creation of these epistemic products depends on the closure of socio-epistemic processes in which social and technical constraints again play a crucial role. As was the case for Wikipedia, when co-producing technological tools, there are technical constraints upon what can be and what cannot be done. Moreover, with respect to certain constraints the distinction between social and technical indeed becomes blurry. For instance, one can conceive the templates we use for reports and documents as technical constraints. However, they only serve as constraints in so far as they are agreed and consented upon, i.e. they may better be conceived as socio-technical constraints.

Finally, as is the case for Wikipedia, social constraints, norms, policies and values play a crucial role for the integration of epistemic products into collective entities. While some of these constraints were quite explicitly discussed and consented upon, others were enacted, modified or abandoned rather implicitly. Examples of the explicit social constraints were the list of deliverables, the time schedule and deadlines for annual reports, the responsibilities of each team, etc. These explicit constraints were articulated in the DOW, the description of work. This document was the common reference point, the solidified agreement to which everyone could refer to when evoking others' responsibilities or for attributing accountabilities. However, work, its distribution, and re-integration do not follow such a schematic description irrespective of how detailed such an account may be. The majority of rules and norms were rather enacted and performed during interactions, while content was collectively co-produced. People created content that others could refer to, – sometimes verbally during meetings or calls, but mostly in written form. Others provided feedback on this – and through this process of providing content and commenting on it, rules and modes of conduct emerged. These rules and modes of conduct are of course embedded in more general academic rules and modes of conduct. Yet they differ among disciplines, countries and institutions and these differences need to be accounted for in interactions. In writing reports, different styles of scholarly writing have to be integrated. Author orders for reports and articles have to be decided upon – with each discipline having different credit-attribution rules. Emails are sent around – half-joking, half-seriously addressing the appropriate time span to reply to emails from partners.

All these social, technical and socio-technical constraints have an impact on the integration processes within the research project. All of them leave an imprint on the epistemic products that results from the processes, which are created and made public for scrutiny by people other than the ones producing them. Once published, these epistemic products have a life of their own: they may be used as input for future socio-epistemic processes or they may be abandoned and never read again. Either way, for them to be created, and especially for them to be declared to be knowledge, they depend upon the closure of socio-epistemic processes. Once closed, they should be an integrated whole, an atomic unit of knowledge, which nonetheless reflects the collective character of its mode of production. Indeed, there was a strong concern that the epistemic products should reflect their history of distributed labor and integration, because one major criterion for the evaluation of the project concerned the quality of the interaction and collaboration between partners itself.

9.2 ES^I and Social Epistemology

Social epistemologists and philosophers of science have addressed ES^I sociality primarily in their analyses on *collaboration* in science. Different authors have stressed different issues related to collaboration in science: How to split, distribute and re-integrate tasks, how to motivate participants and sustain diversity, how to attribute authority and place trust. In the following I address these socio-epistemological considerations in so far as they shed light on the specificities of systems employing of ES^I as opposed to ES^A or ES^S.

9.2.1 Studying Collaboration in Science

Given the ubiquity of collaboration in science and elsewhere and the interest that it has ignited in different academic disciplines, what have philosophers had to say about collaboration? An interesting classification of collaborations in science has been provided by Paul Thagard. In a paper entitled "Collaborative Knowledge", Thagard differentiates four types of scientific collaboration and outlines their respective merits: *Employer/employee, teacher/apprentice, peer-similar* and *peer-different* (Thagard 1997a). I use Thagard's classification and the case studies he draws upon as a reference point to outline crucial topics which have to be addressed within the analysis of socio-technical epistemic systems that employ integrative mechanisms of closure. These topics concern the structure of epistemic tasks, diversity, trust, as well as the relationship between reputation, trustworthiness and authority. But let's first take a closer look at Thagard's four types of collaboration in science.

Employer-employee collaborations are collaborations in which one party can tell the other what to do. That means they are characterized by a difference in status; examples are relationships between laboratory supervisors and technicians or research staff. For *teacher-apprentice collaborations*, such a status difference persists, but the goal here is to train the apprentice to become a full member, i.e. a teacher, herself one day. An example would be the collaboration between researchers and graduate students. For the two kinds of *peer collaborations*, collaborators usually are on the same status level, although differences in status can of course exist. The difference between *peer-similar* and *peer-different* collaborations concerns whether the researchers work in the same or different fields of research. While in *peer-same collaborations* researchers work in similar fields, multi-, inter- and transdisciplinary work would all be instances of *peer-different collaborations* in Thagard's terminology.

Having classified these types of scientific collaborations, Thagard assesses their *epistemic value* with reference to *Alvin Goldman's five standards of epistemic appraisal* (Goldman 1992: 195): reliability, power, fecundity, speed and efficiency. The question is how, with respect to these five standards, epistemic practices are affected by different types of scientific collaboration. In Goldman's social epistemology, the contribution to the achievement of *truth* is the central criteria of his epistemic assessment and he accordingly defines each standard with respect to its contribution to true belief. Thagard (Thagard 1997a), in a similar vein as Longino (Longino 2002c) argues however, that truth might not be the best concept to assess socio-epistemic practices in the sciences. In most cases, whether a theory is true or not, turns out only on the long run. Scientists are therefore rather interested in *publishable results*. If something gets published in a good peer-reviewed journal, he argues, it means it is accepted by an epistemic community and can serve as a proxy for truth, which can only be assessed on the long run – if ever (cf. Solomon 2001, Longino 2002c). In sum, scientists in their daily work rather strive for publishable results than for a rather abstract notion of truth. Accordingly, Thagard modifies Goldman's standards by replacing the reference to truth with a reference to publishable results.

His five standards now read as follows (Thagard 1997a: 247):

> "1. The *reliability* of a practice is measured by the ratio of results to total number of results and errors fostered by the practice;
>
> 2. The *power* of a practice is measured by its ability to help cognizers find results that answer the questions that interest them;
>
> 3. The *fecundity* of a practice is its ability to lead to large numbers of results for many practitioners;
>
> 4. The *speed* of a practice is how quickly it leads to results;
>
> 5. The *efficiency* of a practice is how well it limits the cost of getting results."

Thagard then evaluates his four types of scientific collaboration with respect to these five modified epistemic standards. In *employer-employee collaborations* researchers decide to delegate tasks to technicians or assistants. These tasks they could mostly do themselves, and probably even better.[141] In those cases, delegation might lead to a decrease in reliability. However, this potential loss is made up for by an increase in power, speed and efficiency. The same is true for *teacher-apprentice collaborations*. There might be an additional loss in efficiency, since some of the researcher's time is devoted to training. However, this loss in efficiency is made up for by a gain in fecundity, because once the apprentice is trained, he can produce results of her own (and train other apprentices) (Thagard 1997a).

When applying the standards to *peer-similar* and *peer-different collaborations,* the potential gains and losses become less straightforward. For instance, reliability may just as well be increased than decreased by peer-similar collaboration. Reliability may be increased, if researchers carefully check each others' results and many researchers may even check others work more critically than their own. However, if they spend too much time on cross-checking, they can just as well do the work themselves and potential gains in efficiency are lost. There is also the possibility that they are overconfident in their collaborators and do not cross-check the results in principle or carefully enough. Cases of scientific fraud are often the result of such overconfidence and lack of scrutiny.

Further, the more researchers depend on each other, the more difficult it might be to challenge the other; incompetence or fraud might either be not detected or even obfuscated. As Shapin has already noted for the Gentlemen culture of the Royal Society in the 17th century, withdrawal of trust may come at high (epistemic) costs (Shapin 1994). Cross-checking gets even more difficult in *peer-different collaborations* because often researchers do not even have the skills to check their colleagues' methods or to verify their results. Here trust is even more pronounced (e.g. Hardwig 1991) and reliability might be endangered. However, if *results obtained with different methods converge*, these results are more robust and reliability can be increased through such a triangulation of methods.

Obviously, for peer collaboration, gains as well as losses in the different standards can be expected. So why collaborate, if it is not even sure that this will be epistemically

[141] To my mind, this claim is somewhat problematic. There may be cases in which assistants are better at certain tasks than their supervisors and technicians might be specially trained to do something the supervisor is not able to do. I think this is a reason rather to add another type of scientific collaboration to Thagards list. I would argue that these cases are a mixture between hierarchical differences of the first two types and the differences in specialization characteristic of peer-different collaborations. Hence, they could be labeled *non-peer-different collaborations*.

beneficial? Thagard argues that one important aspect of science got neglected in this list of standards: *explanatory efficacy*. He defines *explanatory efficacy* as the extent to which a practice "[...] contributes to the development of theoretical and experimental results that increase explanatory coherence" (Thagard 1997a: 255). Scientific collaboration and especially *peer-different collaboration*, he argues, can increase explanatory coherence in two ways: *conceptually* by developing new theoretical frameworks by combining conceptual approaches, and *empirically*, by integrating *experimental results* from different fields.

Thagard depicts two case studies from science to show that *peer-different collaboration* in particular can increase his sixth criterion of *explanatory efficacy* profoundly. One example, which he analyzes in a subsequent paper in more detail, concerns *peptic ulcer* (Thagard 1998a; Thagard 1998b). Thagard (Thagard 1997a) argues that until 1981, it was widely held in medicine that peptic ulcer was caused by excess acidity. By 1995, the common view was that it was caused by a bacterium, Helicobacter pylori. How did this "paradigm shift" happen?

According to Thagard (Thagard 1997a) it was a *peer-different collaboration* that initiated this shift. J. Robin Warren, a pathologist noticed the bacteria in biopsies. He could not account for their relevance and therefore sought assistance from a gastroenterologist, Barry Marshall. Together they devised an experiment to empirically test the association between peptic ulcer and the occurrence of the bacteria. When they were also able to show that peptic ulcer could be cured by applying antibiotics the medical community began accepting this new cause of peptic ulcer. To use Solomon's terms (Solomon 2001): premature consensus had to be dissolved before a new consensus could be achieved. In the course of this process, it was not only the two of them, but also medical experts from other fields (e.g. microbiology and pharmacy) who participated in this interdisciplinary collaboration. It was this successful peer-different collaboration, which lead to a new theory on peptic ulcer being widely accepted in the medical community.

9.3 Key Topics

9.3.1 ES^{I} & the Structure of Epistemic Tasks – Or the Two Cultures Revisited

Despite obvious benefits, (interdisciplinary) collaboration is still quite rare in the humanities as compared to the natural and social sciences. Thagard exemplifies this difference by reference to the ratio between single- and multi-author papers in some major journals in the sciences and the humanities. In the *Journal of Philosophy* of 1992 only 4 out of 27 papers are written by two or three authors, while the remaining 23 are written by only one author. By contrast, in the 1992 volume of *Cognitive Psychology* only 4 out of 16 are single-author papers and in the *Physical Review Letter* only 67 out of 558 contributions are single authored (Thagard 1997a: 244f).

Thagard identifies three reasons for this different prevalence of collaboration: *funding, tradition and lack of a natural division of cognitive labor*. A quite mundane reason for this lack of collaboration might simply be that in the humanities researchers less often have funding to employ grad students and that close collaboration is therefore less frequent. Moreover, philosophy for instance, is still considered to be a rather solitary practice of reading and writing, even if intellectual exchange between colleagues clearly is crucial for philosophy. The third point, however, the lack of a natural division of cognitive labor, is most interesting and relates back to Benkler's claims concerning the necessity of modularity and granularity of epistemic tasks (Benkler 2006).

While in many natural and social sciences there are distinct subtasks that have to be conducted in research, such a differentiation is less clear in the humanities. For instance, in experimental research there are different, related, but easily distinguishable steps of research: conceptualizing the experiment, preparing the experiment, finding participants, running the experiment, calculating the results, interpreting the results, writing the reports, presenting the findings, etc. To use Benkler's terminology, one can argue that experimental research is highly *modular.*

Clearly, research in the humanities also consists of different steps and different tasks. I can surely distinguish different epistemic tasks necessary for the completion of this thesis: there were various rounds of reading, excerpting, writing, presenting, discussing, reviewing, etc. However, these tasks are not as easily divisible or distributable over different people, because the final result, the thesis, needs to be particularly coherent, it needs to be *monolithic*. As Benkler has already noted, there are differences with respect to the modularity of different epistemic tasks, depending on their goals. Writing an encyclopedia is different from writing a textbook for high schools, which is why Wikipedia is more successful than Wikibooks (Benkler 2006: 326).

One might wonder whether this means that the humanities are less suited for collaboration just due to the fact that the typical tasks, processes and products are less modular than those one finds in the social and natural sciences. A possible conclusion could then be that one should not request humanities scholars to collaborate. However, given the obvious benefits of collaborative practices, a different consequence might be to question – or even change – the tasks in the humanities. Are different, more modular forms of intellectual work in the humanities feasible? And are they desirable? These are questions any request to increase the amount of large-scale collaboration in the humanities has to address. Yet taking into account the outlined benefits of combining and integrating the competencies and skills of diverse epistemic agents, a rise in collaborative and interdisciplinary projects seems clearly desirable also in the humanities.

9.3.2 ES^{I} & Diversity

In his case study on peptic ulcer, Thagard shows that a multitude of different types of expertise and skills was needed to find the causes of peptic ulcer (Thagard 1998a; Thagard 1998b) . He concludes that for this reason only an interdisciplinary collaboration, a peer-different collaboration, to use his terminology, could possibly have come up with these results (Thagard 1997a). To develop a new theory of peptic ulcer, not only were *new ideas about possible causes* combined that contradicted the mainstream point of view and. It was also necessary to *combine different empirical data* from different disciplines to provide evidence for these theoretical claims. Due to this combination of resources, the explanatory efficacy of this new theory was very powerful, and as a consequence, it was possible to change a widely held belief in the medical community. Referring to yet another example of an interdisciplinary project in cognitive science, in which Thagard participated, he concludes that *peer-different collaboration is epistemically desirable particularly because of its potential to increase explanatory efficacy* (Thagard 1997a).

This conclusion can be read as an advocacy of interdisciplinary research – or, more broadly conceived, as a defense of epistemic diversity in science. Indeed, the epistemic merits of diversity have not only been addressed by social epistemologists and philosophers more generally (e.g. Solomon 2006, Kitcher 1993, Longino 2002c, Bohman 2006, Ober 2006, Kelly 2006), they have also been supported mathematically (Page 2007, Page 2001) and have further been used as an explanation for the success of most, if not all, examples of epistemic social software.

While the benefits of cognitive diversity are widely acknowledged in social epistemology and philosophy of science and beyond, there is little consensus on whether or not diversity has to be reinforced by various interventions. In a paper on "Norms of Epistemic Diversity" Miriam Solomon distinguishes three different socio-epistemological stances towards diversity in science (Solomon 2006). The first one is a laissez-faire view according to which there is enough diversity in science and that for various reasons there should be no intervention into science. Solomon attributes this view amongst others to Philip Kitcher (Kitcher 1993). Solomon contrasts this laissez-faire or invisible hand model with feminist positions that demand explicit encouragement and support of more diversity, e.g (Longino 1990 or Evelyn Fox Keller Keller 1985).

I return to Solomon's analysis of diversity in more detail in Chapter 11, because her perspective is relevant to understand diversity with socio-epistemic systems that function via *selection*. This implies already that diversity can be assessed on different levels and that it functions differently in different types of epistemic sociality. I argue that within systems which depend on integration of epistemic tasks and results, diversity needs to be coupled with shared standards. By contrast, systems that depend on aggregation or selection as mechanisms of closure do not need shared standards to the same extent.

The social epistemologist, whose account I consider most instructive for understanding the relevance of diversity in socio-technical epistemic systems that use integrative mechanisms of closure, is Helen Longino. Longino resolutely defends epistemic diversity and argues that "[a] diversity of perspectives is necessary for vigorous and epistemically effective critical discourse (Longino 2002c: 131). However, as we have learned from her criteria for ideal epistemic communities, such diversity needs to be combined with *shared standards* to lead to epistemically desirable results. For Longino, openness and the diversity of perspectives together with critical interaction are features which make science robust and thus need to be fostered and supported. The primary benefit of diversity in Longino's account consists in its potential to avoid and counter biases. Diversity of perspectives and transformative mutual criticism among epistemic agents with different perspectives is the prerequisite for knowledge to be created, for objectivity to be achieved. To elucidate this interplay of diversity, shared standards and critical interaction, I use a longer quote from Logino:

> "By "social," I have meant "interactive." Certainly, some assumptions and values are shared in any community, but genuine interaction requires diversity among the members. What identifies a given community as a community is a not a set of shared substantive beliefs, but a set of public standards to which community members appeal in critical discursive interactions. These standards may include substantive content, but they also include criteria of evidence and reasoning, methods of investigating. [...] They are [...] regulative elements of critical discourse. They come into existence in the course of the elaboration of such a discourse and are themselves subject to the same kind of critique they regulate. They may be internalized by individuals, but this is a consequence of their status in community interactions and not the other way around. Shared standards permit diversity of beliefs, but unity in their methods of evaluation." (Longino 2002c: 148)

A quite similar ideal of using diversity to counter bias is evoked in Wikipedia's emphasis on pluralism in their description of the "Strengths, weaknesses, and article quality in Wikipedia". They write that "Wikipedia is open to a large contributor base, drawing a large number of editors from diverse backgrounds. This allows Wikipedia to significantly reduce regional and cultural bias found in many other publications, and makes it very difficult for any group to censor and impose bias".[142] As Longino does for science, Wikipedia uses *diversity* and *openness* as arguments for the validity of epistemic practices and systems. Although the openness of Wikipedia can also be a danger, that is acknowledged in their explication of the weaknesses of Wikipedia[143], this openness is essential, because it enables diversity, which is necessary to counter bias. Yet obviously, diversity does not guarantee that bias is avoided and this is why an abundance of norms, guidelines and policies has emerged in Wikipedia as was outlined

[142] http://en.wikipedia.org/wiki/Wikipedia:About#Strengths.2C_weaknesses.2C_and_article_quality_in_Wikipedia [date of access: 17.03.2010]
[143] Please confer also: http://en.wikipedia.org/wiki/Wikipedia:About#Strengths.2C_weaknesses.2C_and_article_quality_in_Wikipedia [date of access: 17.03.2010].

before. These social norms act as constraints, as shared norms of evaluation and conduct to which members of the Wikipedia community can appeal and which get enacted, internalized or transformed via interaction.

Moreover, Wikipedia acknowledges a second function of diversity, namely the possibility to pool cognitive resources. Hence the self-description cited above continues as follows: “A large, diverse editor base also provides access and breadth on subject matter that is otherwise inaccessible or little documented”.[144] Here diversity is considered positive, not because it makes systematic bias less likely, but because it helps to gather necessary competencies, skills and knowledge which may be widely distributed.

To conclude: The dual epistemic function of diversity for systems that use integration as a mechanism of closure consists in pooling of socially distributed epistemic resources as well as in the avoidance or counter of bias. Yet while access to widely distributed knowledge and skills is also characteristic for the other two types of epistemic sociality, the need to *integrate* these competencies implies the needs for shared standards of evaluation and shared social norms of collaboration and interaction. Hence, for diversity to be beneficial in such systems it is not sufficient to simply lump together a bunch of diversely skilled epistemic agents. For socio-technical epistemic systems employing this type of epistemic sociality to successfully produce epistemic content, the individual contributions need to be integrated. This integration crucially depends on shared norms and standards. However temporary they may be, such shared norms and standards are indispensable in such systems. As is shown in the next chapters, this dependence on shared standards is one of the crucial differences between systems depending on integration as opposed to those depending on aggregation or selection as mechanisms of closure.

9.3.3 ES^I: Trust

The relevance of diversity and task structures is not the only topic raised in Thagard’s account of scientific collaborations. He also makes reference to Hardwig’s work on the central role of trust in science and for knowledge more generally. Since Hardwig’s

[144] http://en.wikipedia.org/wiki/Wikipedia:About#Strengths.2C_weaknesses.2C_and_article_quality_in_Wikipedia [date of access: 17.03.2010]

position has been outlined before, a quote should suffice to summarize his claims. Hardwig states:

“Modern knowers cannot be independent and self-reliant, not even in their own fields of specialization. In most disciplines, those who do not trust cannot know; those who do not trust cannot have the best evidence for their beliefs. In an important sense, then, trust is often epistemologically even more basic than empirical data or logical arguments: the data and the argument are available only through trust. If the metaphor of foundation is still useful, the trustworthiness of members of epistemic communities is the ultimate foundation for much of our knowledge” (Hardwig 1991: 693).

This quote indicates the centrality of trust for knowledge: we have to trust other epistemic agents as well as epistemic content to use it as input for further epistemic processing. For epistemic content to be considered trustworthy, we further have to trust non-human epistemic agents as well as the processes involved in the creation of this epistemic content. Trust then pervades all phases and instances of socio-technical epistemic systems and processes. The following picture visualizes this pervasiveness of trust in socio-technical epistemic systems that employ integrative mechanisms of closure. As will become obvious in the following chapters, the loci of trust differ between systems that employ different mechanisms of closure.

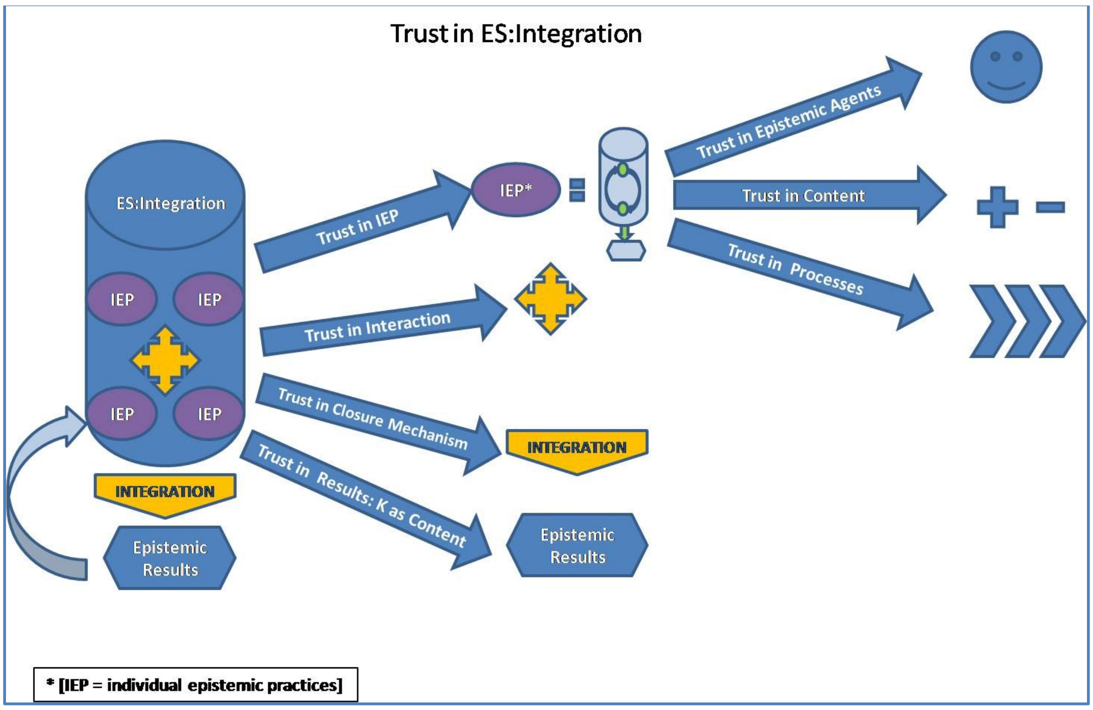

Figure 14: Trust in ES:Integration

Referring back to Longino's differentiation of different senses of knowledge, I take it as a central premise that, to understand science as well as open-ended epistemic practices on the Web, we have to focus on the interaction between *knowledge as content* and *knowledge-producing practices* (Longino 2002c: 77ff). Knowledge as content is the input just as much as the output of knowledge-producing practices. In order to use knowledge as input for epistemic practices, we have to trust this knowledge. This implies trust in the people and mechanisms that have produced this knowledge. We cannot possibly test every premise before accepting it. We have to stop scrutinizing somewhere and trust. We trust people, we trust methods, we trust our senses – without trust, knowledge is impossible. In this sense trust is fundamental for knowledge: without trusting some knowledge by using it as input for our epistemic practices, we cannot engage in the process of creating knowledge. At the very least we have to trust our senses or our reasoning, yet in the majority of epistemic situations, we have to trust in much more profound ways: other people, instruments, methods, etc. Science and the Web in particular are environments in which epistemic processes are less likely to be finalized by direct recognition. Indeed, scientific knowledge is always provisional, partial and plural, as Helen Longino has convincingly argued (Longino 2002: 207). I do think that the same is true for knowledge as content and knowledge-producing practices on the Web.

Thus trust seems to be indispensable for knowledge creation in science and beyond. And indeed, Thagard also agrees with Hardwig (Hardwig 1985; Hardwig 1991) that "we very much epistemically dependent on one another: much of what each of us professes to know depends on information that we have acquired from others that we trust" (Thagard 1997a: 241). However, he immediately alerts us to the *dangers of trust* for science when he continues: "The cost of epistemic dependence of the sort especially notable between collaborators is that mistakes can enter and propagate within the system because of collaborators who are inept or corrupt. Hence collaboration between equals may decrease as well as increase reliability" (Thagard 1997a: 241). Here, Thagard hints at the delicate equilibrium of trust which is needed for collaborations to score well with respect to the five standards of epistemic appraisal. Too much trust – or blind trust – may reduce reliability and hence have detrimental epistemic effects by leaving questionable results and methods unquestioned. Too little trust may lead to losses in efficiency and speed because agents would waste too much time in checking others' results. Not to mention the more atmospheric problems that will occur when distrust becomes obvious. If trust is suspended altogether, collaborations become impossible.

This ambivalence is inherent in the concept of trust itself. In her article on trust in the Stanford Encyclopedia of Philosophy, Carolyn McLeod states that “trust is an attitude that we have towards people we hope will be trustworthy” (Mcleod 2006). In line with my previous arguments, I would add that trust is an attitude that we not only have towards people, but also to knowledge as content, to mechanisms, processes, to non-human epistemic agents more broadly conceived. Yet as the word *hope* indicates we cannot be sure that what we trust really will prove to have been trustworthy in the end. Trust is not being certain. Trust by definition contains the risk of being let down, the risk of having placed trust wrongly – and one of the crucial philosophical questions about trust is under what conditions trust is warranted.

Relating back to Hardwig’s original paper, trust involves a reliance on others’ *competence* and their *honesty*. Trust then is warranted if we identify those epistemic agents who are competent and honest, if we can distinguish the honest from the dishonest, the competent from the incompetent. This identification implies a differential epistemic and moral assessment of epistemic agents.[145] However, this assessment comes with several problems. First of all, assessing someone else’s competency is difficult in fields in which we are not competent ourselves. Assessing someone’s honesty also is a difficult task if we do not know what is at stakes for him. Second, even if we have the means to assess someone’s competency, some of the benefits of collaboration (with respect to efficiency and speed) would be diminished if we spent too much time cross-checking others. Thus, in cases where we lack the means to assess someone’s competency or honesty directly or if we want to save time, we make use of different *proxies of trustworthiness*.

As outlined in Chapter 6, Shapin (Shapin 1994) offers one rule of a thumb to decide upon the *honesty*, namely *disinterestedness*. While analyzing the development of English experimental philosophy as a predecessor of experimental science, Shapin argues that a good proxy to find someone who does not lie is to find someone who is

[145] Hardwig (1991) introduces a third criterion to assess someone’s trustworthiness: adequate epistemic self-assessment, i.e. the ability to assess one’s own level of competence regarding the issue at hand. Since adequate epistemic self-assessment is a *second-order competence*, which indicates the limits of one’s competence, I for the moment subsume it under my considerations on the assessment of competence.

disinterested, someone who is free. And in Seventeenth-century England, those who were considered to be free were the gentlemen, because their economic position allowed them not to work. Economic freedom was understood as a prerequisite for moral freedom. It was assumed that "[g]entlemen were truth-tellers because nothing could work upon them that would induce them to be otherwise" (Shapin 1994: 84). Indeed, Shapin argues, "[...] it was the disinterestedness of the English gentleman's situation that was most importantly identified as the basis of his truth-telling" (Shapin 1994: 83).

Philipp Kitcher (Kitcher 1993) is another author who has stressed the crucial role of trust for collaboration in science and acknowledges that trust not only needs to be placed in people, but also in instruments, techniques and procedures (Kitcher 1993: 308)). The major part of his analyses, however, concerns the assessments of peers. In contrast to Shapin, however, Kitcher does not focus on an assessment of others' honesty, but on an estimation of their *competency*. The guiding questions for him are: "Whom should one trust? When should one trust others more than oneself? When is it worth risking the errors that others might make?" (Kitcher 1993: 304). Kitcher emphasizes the differential assignment of trust to different epistemic agents – oneself included. This differential assignment of trust is based on an *estimation* of the competence of epistemic agents based on different proxies. That is, when competence cannot be directly assessed, people use different indicators to predict the competence of others. Sometimes it may be possible to rely on the evaluation of *past* performances. If that is not possible, one may use someone's social position, her institutional background, etc. as indicators for her competence. Kitcher uses the term *earned authority* to refer to the former and *unearned authority* to refer to the latter (Kitcher 1993: 315).

This leads us to the relationship between trust and authority. In order to know, we have to trust. Yet trust implies the risk of misplacing trust, of placing trust in agents who are incompetent or dishonest, in instruments, procedures and information, which are not valid. Trust is not certainty. But in order to be epistemically responsible knowers, we have to strive for placing trust adequately. And to do this, we have to assess epistemic agents and processes. For the assessment of human epistemic agents this implies that where we cannot assess their competence and honesty directly, we may have to rely on proxies of trustworthiness. These proxies can and often are reputational cues, indicators of past performance as well as various social cues – Kitcher's unearned and earned forms of authority. However, I would argue that these cues are not types of authority themselves, but indicators of how to attribute authority to others. Authority in this sense implies the assignment of *power* based on the making of differences between epistemic agents based on various – more or less valid – proxies of epistemic trustworthiness. Indeed, power functions twofold in this process. On the one hand existing differences in societal status may have an effect on the attribution of epistemic authority to epistemic agents as feminists have shown (e.g. Alcoff 2001). On the other hand, differences in power are also the *result* of such differential attributions of authority and epistemic trustworthiness. I return to questions of authority, trustworthiness and reputation in the next section.

Before, let me conclude these explanations on the role of trust in socio-technical epistemic systems and processes. In order to know we have to put some trust in knowledge, we can never be certain about all the premises we use in our knowledge-producing practices. However, since trust is not certainty, it involves the risk of being let down. And as responsible knowers, we have to be aware of this risk. We may have a default to trust, but also a duty to watch out for signs of dishonesty and incompetence. Therefore the differential assessment of peers regarding their competency and honesty is a major issue for analyzing socio-technical epistemic systems in which multiple epistemic agents are involved in epistemic practices. And this differential assessment leads us to another cluster of issues: the relationship between authority, reputation and trustworthiness.

9.3.4 ES^{I}: Authority, Reputation, Trustworthiness: Making Differences that Matter

While in the last section it has been noted that trust can and has to be placed in people as much as in non-human epistemic agents, in techniques, processes and mechanisms, I focus in this section on human epistemic agents. I analyze and show how we use reputational cues to assess the trustworthiness of epistemic agents, how we attribute authority based on this assessment. This section is on the power of making differences that matter between epistemic agents.

Coming back to Thagard's analysis of the respective merits of different types collaborations in science, it becomes obvious that the main criterion for distinguishing these different types consisted in the *respective authority of epistemic agents*. The distinction between peers and non-peers is essentially a distinction based on the amount of authority ascribed to different epistemic agents. This differential attribution of authority is then used as a criterion for attributing differential amounts of trustworthiness to epistemic agents.

As noted before, Philip Kitcher examines the role of authority in his formal analyses of the organization of cognitive labor (Kitcher 1993). He starts with the assertion that collaboration is advantageous, because it allows scientists to achieve some goals faster and makes other larger tasks, tasks that would extend the lifespan of one person, feasible in the first place. However, in those tasks that could be conducted individually, the decision to collaborate depends not only on the gain in time, but also on the assumed error rates of doing the task yourself versus delegating it to someone else. That is, gains in speed have to be weighed against potential losses in reliability. This estimation of error rates assumes that scientists can – and do – assess their peers' authority, i.e. they estimate each others' reliability to decide upon collaborations and deference to authority. Kitcher outlines different options scientists have in order to assess their peers. One of the crucial distinctions is between *earned* and *unearned authority*. While *unearned authority* for Kitcher refers to scientists' social position or their institutional background, whereas the quality of their papers and the journals they have published in refers to their *earned authority*. Moreover, scientists can assess their earned authority via direct or indirect calibration. *Direct calibration* refers to assessing epistemic claims oneself, e.g. by reading an article by someone to assess her competence, whereas *indirect calibration* refers to assessing her competence by taking into account how others have judged her article, e.g. via ratings, recommendations, etc.

Kitcher crucially notes that the idea that non-epistemic, social factors play no role in the attribution of authority is a "traditional epistemologist's utopia" (Kitcher 1993: 318). *Alliances, prestige effects and backscratching* are the topics that he outlines to show the extent to which social factors and relations do influence or bias the attribution of authority. The crucial criterion to assess such practices for Kitcher however remains purely epistemic: to what extent does deference to authority save time and resources and to what extent does it lead to stagnation and hence block innovation and scientific progress.

However, many feminist scholars have stressed that such socially induced biases in attributing authority are epistemically *and* ethically highly problematic (e.g. Fricker 2007, Code 1987, Scheman 2001, Campbell 2001, Alcoff 2001). To what extent is the social identity of epistemic agents relevant for the assessment of their capacities as a knower? Which proxies are valid as indicators of her epistemic trustworthiness? Which reputational cues should be used for ascribing authority? And how do we detect and avoid the use of invalid proxies? The use of reputational cues to assess epistemic trustworthiness and to attribute authority has epistemic advantages, especially in the absence or - as is particularly true in the case of the Web - in the overabundance of information. Yet it also comes with epistemic and ethical dangers.

First of all, the use of reputation to assess content can be epistemically beneficial while being morally questionable. Once social information is taken into account to rate the quality of content, the door is open for social biases, prejudices and discrimination, which are as prevalent in science and on the Web as in the societies that have developed and maintained these socio-technical epistemic systems. These problems are not new and have long been identified for science and other epistemic fields by feminist epistemologists. In addition to raising awareness about these problems, various scholars have also developed tools and strategies to counter these epistemic injustices (Fricker 2007, Scheman 2001, Alcoff 2001). Miranda Fricker, for instance, distinguishes between testimonial and hermeneutic injustices as two instances in which someone is wronged in his capacity as a knower based on his social position. According to her "testimonial injustice occurs when prejudice causes a hearer to give a deflated level of credibility to a speaker's word, whereas hermeneutic injustice "[...] occurs at a prior stage, when a gap in collective interpretative resources puts someone at an unfair disadvantage when it comes to making sense of their social experience" (Fricker 2007) 1). Clearly, both forms of injustice are easily conceivable when reputational cues and their epistemic usage are not critically reflected upon and kept open for constant scrutiny and revision.

The second problem concerns the limits of the epistemic usefulness of reputational information itself and can be further split into two types of concerns. The first question is how you calculate the reputation of someone else in the first place, resp. which proxies you use. Do you use the person's academic development, his institutional background, some form of communal evaluation, such as ratings or recommendations that he has received from other people as a cue to assess someone's reputation? Do you rely on your own experience with her? Or some indicator of the quality of her former research? On her track record of different academic achievements? Her H-index or impact factor? Which of these proxies are valid and which are not? The second crucial question concerns the stability of reputation, i.e. the way you deal with evidence that supports or contradicts your view on the reputation of others. When, under which conditions, and up to which degree of counter-evidence are you warranted in keeping your reputation value for someone or something? If we want to explore the utility of reputation for epistemic purposes, we have to analyze the potentials and possible dangers very carefully. That reputation is used to assess information and epistemic claims goes without saying – and it comes with benefits as much as with problems. So the question should be less how to avoid using reputation cues for epistemic purpose, but rather how to use them wisely.

9.4 Implications for the Analysis and Design of Socio-Technical Epistemic Systems

9.4.1 Trust & Authority in Wikipedia

Given the relevance of trust for knowledge creation in everyday life and in science, it should not come as a surprise that trusting to know is crucial on the Web as well. The Web is an enormous conglomeration of information of varying quality. If we pursue epistemic goals, it is therefore crucial to extract valuable information from the overabundance of existing information. And again: in many cases we have to trust because we cannot check everything for ourselves. In the following I apply some of the insights obtained before to Wikipedia.

65 million people use Wikipedia each month. If there is one conclusion to be drawn from this number, it is that many people use Wikipedia for epistemic purposes. Yet can you trust Wikipedia? And if so for what reasons? Different answers have been given to these questions (Magnus 2009, Tollefsen 2009, Wray 2009) and some of Wikipedia's critics keep insisting that one should not trust it at all (Waters 2007, Keen 2008). Indeed, many of the proxies that we usually use to assess the trustworthiness of information are missing: due to its openness, people can edit who would not be considered experts on the topic. Due to the anonymity of contributions we can neither assess the competency nor the honesty of contributors. The reason for trusting Wikipedia must thus be a

different one than trusting a person. I argue that the reason why people trust the content of Wikipedia is that they *trust the processes of Wikipedia*. It is a form of *procedural trust*, not a trust in persons.

If people know how Wikipedia works and trust it, then they trust it, because they attribute some epistemic authority, some trustworthiness to the process by which Wikipedia generates information. They do not trust specific persons; they trust Wikipedia as a system that is based on a distinct process of information provision. This process is characterized by a multitude of users that can edit and change information immediately, combined with a mechanism to undo revisions easily and quickly as well as a system for tracking these changes and making them visible. It is a system enforcing open access with minimal entry barriers, combined with mechanism of making editing patterns transparent.

The rationale behind this openness of Wikipedia is scale. More people can provide more information faster. However, a second point is crucial: the more people scrutinize and can easily change content, the less likely errors and biases remain undetected. Please note that there is no guarantee that an error or bias gets detected. It is only *more likely,* if a multitude of diverse people participate in this process. This is a probabilistic statement.

Wikipedia provides *a lot of* information on many topics due to the multitude of volunteers, due to its diversity and the ability to pool abilities, competencies and socially distributed knowledge and information. Yet if Wikipedia did not provided *good* information, it would not be used to the extent it is used. On average, Wikipedia works well as a source of information. However, the problem lies exactly in the term *on average*. Since Wikipedia is a dynamic system, information can be changed by anyone anytime. Hence, there is no guarantee that the information provided at the moment one accesses Wikipedia is correct or a blatant lie. While Wikipedia might be trusted *as a system* because of certain characteristics, this overall trustworthiness does not help us to assess the trustworthiness of a *specific claim* in Wikipedia. This is the crux of statistical reasoning: it works well on average, but we can never be *certain* about a specific instantiation. To improve this assessment of specific claims, various tools have been developed to empower users to make more informed decisions about which information to trust and where to be skeptical. In the following I depict two of them briefly, the WikiScanner and the WikiDashboard. While both of them empower users to assess the trustworthiness of information, they function very differently – and require different forms of trust themselves. I conclude this section with sketching yet another example of a tool that can empower users and make them more careful and responsible knowers: traffic lights of trustworthiness.

WikiScanner: Which IP-Address Does Not Belong to a Gentleman?

In 2007, a tool called WikiScanner[146] has received quite a lot of attention and media coverage. WikiScanner is a search tool that traces IP addresses of those who change Wikipedia entries anonymously. The tool was developed by Virgil Griffith, a graduate student at CalTech. Griffith says that the inspiration for this tool has been the revelation that the offices of Congress members had been editing their own Wikipedia entries (Borland 2007). He wanted to find out whether other companies and organizations also edit entries in ways that served their interests. By tracking the IP addresses of anonymous editors, this tool unveiled that this was indeed the case. Numerous organizations edited Wikipedia articles to serve their interests. On his website, Griffith concludes: "Overall - especially for non-controversial topics - Wikipedia seems to work. For controversial topics, Wikipedia can be made more reliable through techniques like this one."[147] What is the epistemic and ethical utility of a tool like WikiScanner? Referring back to the insights obtained from Shapin, one may argue that WikiScanner helps to find out which IP-addresses do not belong to gentlemen Shapin 1994. Disinterestedness is an important proxy for honesty and the flipside of the coin is that if we can reasonably assume bias, because those who edited an entry do have an interest in it, we should re-assess the amount of trust we place in this particular information accordingly.

WikiDashboard: Using Editing Patterns to Assess Trustworthiness

WikiDashboard is another tool which aims at raising the trustworthiness of Wikipedia by making things transparent. Yet the rationale is quite different. Instead of revealing the identity of anonymous contributors, WikiDashboard "[...] visualizes the social dynamics and editing patterns of every article and editor of Wikipedia" (Chi, Suh et al. 2008). The basic idea behind the development of this tool is that the fact that anyone can edit any Wikipedia article is not necessarily a threat to reliability. Although WikiScanner reveals

[146] http://wikiscanner.virgil.gr/
[147] FAQ of WikiScanner: http://virgil.gr/31

that people sometimes consciously and purposefully introduce error into Wikipedia and by doing this exploit Wikipedia's openness, the same openness can also be a source for its reliability. Using science as a point of reference, Suh et al. (Suh, Chi et al. 2008) argue that it is precisely the possibility to put ideas into discussion, to examine and challenge each other's claims, that is crucial for knowledge generation in science and that similar processes also occur on the Web. Reliability and growth of knowledge might thus be advanced by discussions and mutual criticism in combination with practices that increase social transparency, such as attribution, indication of past performance and provision of sources.

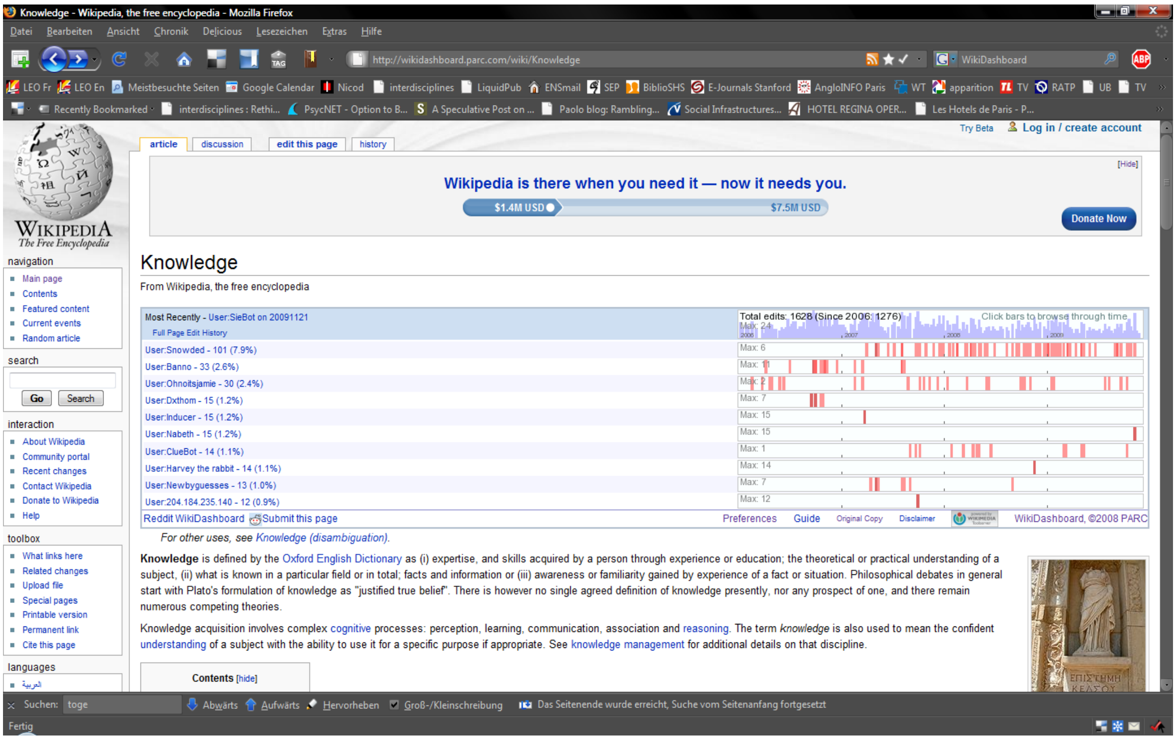

Figure 15: Screenshot of WikiDashboard for the article "Knowledge"[148]

WikiDashboards exist for users and for articles.[149] Thus the editing activity of a specific user or a specific article is visualized and can be used as a cue for assessing the trustworthiness of an article at a given moment or as a proxy for the trustworthiness of a user. The WikiDashboard embedded within each article of Wikipedia is intended to make the user aware of interesting editing patterns he or she might otherwise not notice. In the case of articles, examples would be sudden bursts of editorial activity due to recent events. WikiDashboards on user sites might indicate the user's specific editing

[148] http://wikidashboard.parc.com/wiki/Knowledge [date of access: 28.11.2009]
[149] Please confer the screenshots I have provided for both cases.

habits as well as the range and variety of topics she contributes to. Such editing patterns could thus be used as proxies to assess the user's competency and honesty. If he has edited many pages with similar content, this may be a proxy for his competence on the topic. If his editing patterns reveal frequent repetitive revisions, he may have been involved in flame wars, which in turn may be a proxy for his dishonesty.

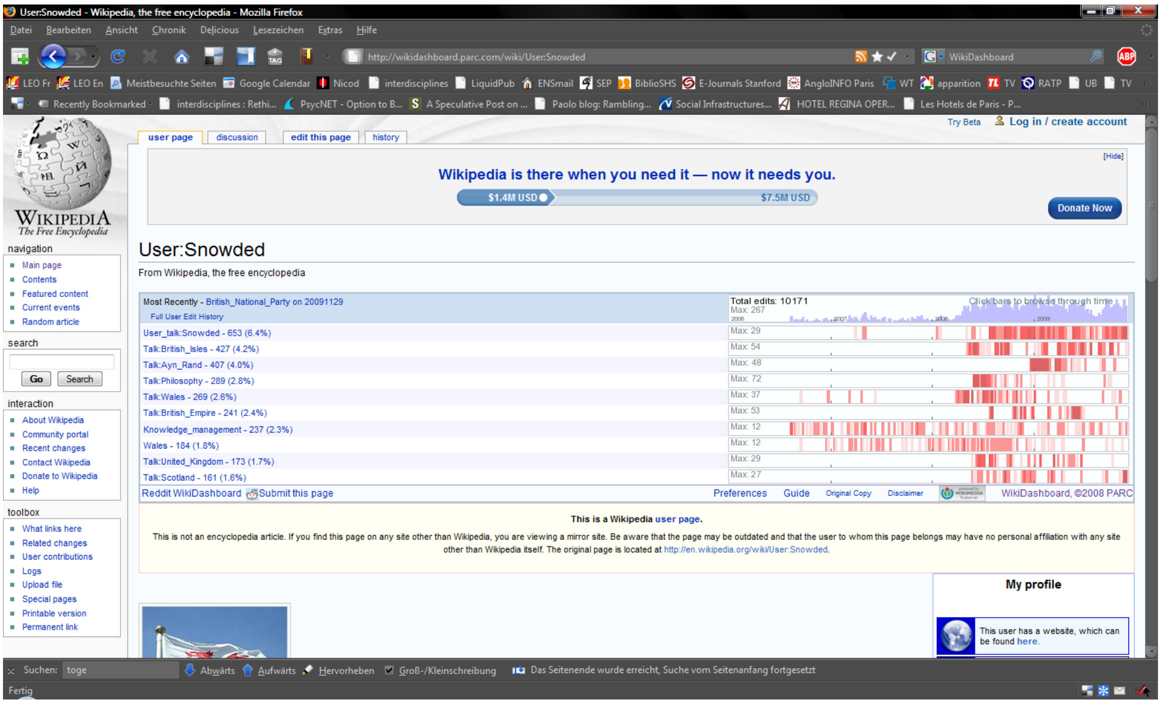

Figure 16: Screenshot of WikiDashboard for the User "Snowded"[150]

To conclude, by unveiling the evolvement of articles and the role and amount of conflict therein, WikiDahsboard aims at raising the social transparency and accountability of Wikipedia. More specifically, the *overall trustworthiness of Wikipedia* is supposed to be raised by an increase of transparency on the level of individual articles and users. Thus, WikiDashboard may *decrease the trust* of users in an article by revealing an editing pattern that is indicative of a flame war and therefore warns the user *not to trust* the information provided. By doing this, WikiDashboard empowers the user in giving him *reasons to distrust*. Yet the existence of such a tool *increases the overall trustworthiness* of Wikipedia, because it amends the information with an indicator of its trustworthiness. To use a statistical comparison: it is as if Wikipedia would not only deliver a measure, but also the confidence interval. One problem with WikiDashboard concerns the interpretation of these visualized editing patterns: to what extent are they useful to assess

150 http://wikidashboard.parc.com/wiki/User:Snowded [date of access: 28.11.2009]

the trustworthiness of users and pages? After all, the user involved in the flame war may just as well be the one who introduces error as the one who devotes time to correct it. The editing pattern alone may not be able to tell the difference. The second problem concerns the accessibility of WikiDashboard. The interface of WikiDashboard it quite complex and some users may refrain from using it, because they do not understand it. This leads us to the relationship between simplicity and complexity in making things transparent.

Transparency, Simplicity and Traffic Lights of Trustworthiness

In the case of WikiScanner and WikiDashboard transparency is used to support critical assessment and attribution of trustworthiness. Rendering the sources and editing patterns visible enables rational assessment of information provided on Wikipedia. However, transparency is often a matter of degree and how transparent something is depends on someone's abilities. The question therefore is: what is how transparent for whom?

For instance, the discussion and history pages of Wikipedia, which serve as input data for Chi et al.'s tool (Chi, Suh et al. 2008) are in principle already accessible to all users of Wikipedia. However, many people do not look at these pages. Maybe because it is too much effort; maybe because they do not want to be bothered; maybe because they do not understand the interface or how information about revisions might be useful to assess the quality of information. It is especially for this very last fraction of users that the WikiDashboard might be interesting. However, even some of those from the 'I don't care'-category might be turned into 'Well, now-I'm-interested', if the indication of trustworthiness is a salient feature of the website and if it had proven useful to them in their first trials.

WikiDashboard is not simply another Web2.0 widget, but a tool for empowerment, a tool that raises awareness about the functioning - and possible malfunctioning - of one of the major information sources on the Web. However, I also think that its interface is quite complex and difficult to interpret. Users may still feel overwhelmed by the huge amount of data provided even in this aggregated format and therefore they might not use it. An interesting question therefore is whether there might be even simpler tools that still raise the epistemological and ethical reflexivity of many differently skilled and interested users.

Let's assume we have an algorithm that translates the editing patterns into the three different signals of a traffic light: Depending on the editing patters, the traffic light embedded on top of each Wikipedia site would be in red, yellow or green. Wouldn't this make you stop and wonder for a second what this signal is trying to tell you? Whether you should treat the information of the article differently depending on whether the lights are on red or green? The traffic lights are simply a much more condensed and intuitive format than the rather complex interface of the WikiDashboard. The traffic light signal would be a dynamic, automatically generated indicator of the controversies which preceded the temporal fixation of the Wikipedia article the moment you read it. This indicator can then be used as a proxy for the trustworthiness of the current state of the article. That this indicator is dynamic and automatically generated based on the editing patterns is crucial for assessing the trustworthiness of any dynamic website. If the traffic lights cannot potentially change with each revision of the article, they would soon be rendered unreliable and thus useless as indicators of trustworthiness of a potentially constantly changing article. Yet this is exactly what users are interested in: a quick assessment of the quality of the article they see at the moment they see it.

However, simplicity comes with a price. If we use the traffic lights as proxies to assess the trustworthiness of information provided on Wikipedia, we only shift the locus of trust: Instead of trusting the *content* on Wikipedia, we now have to trust the *mechanism assessing its trustworthiness*. That is, instead of trusting the mechanism by which information is created on Wikipedia, we now have to trust the algorithm that simplifies a multitude of complex editing patterns into a tripartite signal. And as we will see in the next chapter, decisions embedded into algorithms are even less subject to critical scrutiny than Wikipedia articles – because they are less visible. Yet once we realize the epistemological and ethical relevance of seemingly minor programming decisions, it becomes obvious that we have to maintain and possibly even increase a critical stance when delegating the assessment of trustworthiness to algorithms. And again, a crucial prerequisite of such critical assessment is transparency.

9.4.2 Social Norms, Diversity & Authority in Socio-Technical Epistemic Systems

In this section I argue for the fertility of using Helen Longino's *Critical Contextual Empiricism* to analyze and develop socio-technical epistemic systems. I start with a brief reminder of her approach before using her four social norms for social knowledge to assess socio-epistemic practices on the Web as well as the question of how scientific publishing may be changed through the use of Web2.0.

As noted before, Helen Longino's *Critical Contextual Empiricism* aims at reconciling philosophical accounts of science with insights from the social studies of science by dissolving a dichotomous understanding of the social and the rational. Departing from a decided feminist position, she acknowledges that the *sociality of* knowledge in its three senses allows for biases and distortions in science and scientific knowledge (Longino 2002c, Longino 2005, Longino 1996, Longino and Lennon 1997). However, sociality is not only a biasing factor – it can also deliver the means to counter such biases. The role of a social epistemologist therefore is to show "[...] how the social dimensions of cognition have resources for the correction of those epistemically undermining possibilities" (Longino 2002c: 205). And these resources for correction lie within a *combination of epistemic diversity with shared standards of evaluation.*

In contrast to Philip Kitcher (Kitcher 1993), Longino does not consider science to sufficiently employ and exploit beneficial social mechanisms to counter biases. Hence, in order to make sure that "[t]he social is not a corrupting but a validating element in knowledge" (Longino 2002c: 122), Longino proposes four social norms for social knowledge. These norms were portrayed before, but given their relevance for the following analyses I quote them again (cf. Longino 2002c: 129ff):

> "1. Venues. There must be publicly recognized forums for the criticism of evidence, of methods, and of assumptions and reasoning. […]
>
> 2. Uptake. There must be uptake of criticism. The community must not merely tolerate dissent, but its beliefs and theories must change over time in response to the critical discourse taking place within it. […] Uptake is what makes criticism part of a constructive and justificatory practice. […]
>
> 3. Public Standards. […] Participants in a dialogue must share some referring terms, some principles of inference, and some values or aims to be served by the shared activity of discursive interaction. […] A community's standards are themselves subordinated to its overall cognitive aims […] Finally, standards are not a static set but may themselves be criticized and transformed […] There is no particular act of adopting or establishing standards. […]
>
> 4. Tempered Equality. Finally, communities must be characterized by equality of intellectual authority."

These four norms are more than just ethical or political nice-to-have's. If knowledge involves justification, if knowledge as an epistemological status has to be attributed to a claim by a group of people, then knowledge fundamentally rests on communities and the social mechanisms and procedures they employ. As Longino states, "[w]hile intersubjective interaction is a necessary feature of scientific cognition, not just any form of interaction will do."(Longino 1994: 144). Rather, interaction has to be effective to yield in knowledge. And the four norms exemplify those criteria that have to be met for effective criticism, for checks and balances to take place. *It is only through such effective criticism that the subjective can be transformed into the objective*, "[...] not by canonizing one subjectivity over others, but by assuring that what is ratified as knowledge has survived criticism from multiple points of view" (Longino 2002c: 129). Thus effective criticism is a *prerequisite* of knowledge, and not just an addendum.

Longino considers her four norms as recommendations for the *design and constitution of scientific communities*. An ideal scientific community would fulfill all these four standards to a maximum. In such a community all hypotheses and theories would be subject to the broadest range of valid criticism by fostering epistemic diversity, by encouraging dissenting voices instead of silencing them, and by listening and responding to criticism in ways that will further increase knowledge. The effect of such transformative criticism will be that the "[…] theories and hypotheses accepted in the community will not incorporate the idiosyncratic biases (heuristic or social) of an individual or subgroup" (Longino 2002c: 134). Given that these norms represent an ideal epistemic community, real epistemic communities will hardly ever fulfill all those norms completely, the quality of epistemic communities rather comes in degrees (Longino 2002c: 134). However, the fact that ideals might never be attained does not diminish their value as ideals to strive for. After all, the possibility that truth may not be attained has never hindered people to strive for it.

The utility of Longino's norms for this thesis should be evident. If those norms can guide and amend scientific knowledge creation by giving concrete recommendations of how to design interaction in scientific communities, then the same norms may well be used to design and implement other socio-technical assemblages that serve epistemic purposes.

If we were to design socio-technical systems for epistemic purposes that aim at fulfilling Longino's standards, how should we go about? In other words: what are the characteristics of systems that support transformative criticism and which features of Web tools might be useful? As should be obvious from my remarks in Chapter 3 on Science and Technology Studies (STS), I do not assume that one can design socio-technical systems like machines that function according to plan. According to a more *performative view*, socio-technical assemblages evolve over time and change with practice. However, by designing systems just as much as by designing technical artifacts, one sets parameters that imply certain *limits and affordances,* that make some forms of usage more *likely* than others. This also means that while I might design systems to support transformative criticism, there is no guarantee whatsoever that my system will actually fulfill this task. Even if I set the parameters right, the success or failure of my system – with respect to the goal of being a system that supports transformative criticism - will depend on other factors, such as the motivation of the users, etc. But let's take a look at how one might set the parameters so as to increase the likelihood that a system will support transformative criticism.

The Need for Publicly Recognized Venues for Criticism

The first norm concerns the *need for publicly recognized venues for criticism*. These venues could of course be physical places as well. Given the topic of this thesis, however, I focus on new possibilities offered on the Web. As noted before, the demand for publicly recognized venues for criticism actually implies two demands: not only do we need to devote space to criticism, criticism and the venues should also be publicly recognized.

The issue of space is quite easily solved on the Web. In contrast to print publications, where space was a precious resource, space is not a substantial expense factor on the Web. Gaining *attention* rather is the crucial factor (Franck 1996; Goldhaber 1996a; Goldhaber 1996b, Nielsen 1999). And since it has been argued that science itself also functions according to *attention economics*, there might be one more commonality between scientific practices and epistemic practices on the Web (Kerres, Euler et al. 2005). If we use the example of scientific publishing, economically nothing would speak against devoting as much space to original research as to critical responses to it.

The problem of recognition is harder to tackle. For a long time, original research has been considered to be much more important than critical analyses of other people's work. Just think about the perceived value difference of publishing a study versus a book review. Clearly, appraisal and refutation is in principle an essential part of research. Looking at scientific papers, there is always reference to those works one is building his research upon, as well as those one strives to refute. Nonetheless, such reference may remain superficial and direct replications of work are much rarer than one might expect given the alleged importance of replicability in science. One reason for this is that replication is not much valued in science. Maybe this is the case because the initial act of creativity is missing when an experiment is replicated. Not only do replications hardly receive funding, the chances that those results will be published are also lower. Given the space restrictions of journals, an editorial policy that favors original research might make sense. If one recapitulates Goldman's recommendations for editors, the newsworthiness of articles *should* play a role for selection (Goldman 2003: 263ff). Put more bluntly, replicating something that has been done has no newsworthiness. The only exception are replications of highly surprising original research. Hence, the reputation of replications is low. If one sticks to the criterion of newsworthiness, then maybe their value indeed is low. If a replication of the experimental design leads to the same results this results supports the original finding, but adds nothing new. However, this judgment can only be made *after* the replication has shown a difference and the low reputation of replications might diminish the likelihood of replications in the first place.

A second related problem concerns results which are not significant. Non-significant results are usually also not published or at least are less likely published, although they indeed are results. The result is negative, i.e. there is no difference before and after an intervention or there is no difference between two different treatments, etc. Given the space constraints of paper journals studies that have positive results, i.e. that can show significant differences, are usually favored. This selection practice however, has lead to a significant bias in published results. Assume you are interested in the effectiveness of two different methods of teaching French. You can conduct a meta-analysis of all published articles that compare these two methods and the result is that in 60 out of 80 studies method A has lead to better results. Let's assume further that method A is only moderately more expensive than method B. Than this result would probably lead you to the conclusion that method A should be preferred. Imagine however, that these 80 studies are only the significant tip of an iceberg of 1000 unpublished studies, which showed no difference between the two methods. How would the knowledge about this relation between published and unpublished material affect your judgment?

One possibility to counter this problem of bias would be to publish all studies irrespective of whether they have obtained significant results as long as they meet the standards of their disciplines. Recently, the journal *Rejecta Mathematica* has been launched that publishes papers that have been rejected from peer-reviewed journals. The reasons behind the introduction of such a "second-chance-journal" for mathematical papers are well worth quoting at length. The initiators of *Rejecta Mathematica* argue that the value of publishing rejected papers lies in the following aspects:

> "“mapping the blind alleys of science”: papers containing negative results can warn others against futile directions;
>
> “reinventing the wheel: papers accidentally rederiving a known result may contain new insight or ideas;
>
> “squaring the circle": papers discovered to contain a serious technical flaw may nevertheless contain information or ideas of interest;
>
> “applications of cold fusion": papers based on a controversial premise may contain ideas applicable in more traditional settings;
>
> “misunderstood genius”: other papers may simply have no natural home among existing journals.”"[151]

Other journals exist that publish negative results only.[152] A possible downside of publishing all rejected material may lie in the risk of *increasing the noise*. The more papers get published, the more resources we need to process to find what we are searching for. Whether this fear of information overload is a reason enough to refrain from such a change in publishing would have to be debated. However, another possibility would be to develop a database in which all results – and only the results of studies which meet the methodological standards of their discipline - are saved. Such a database of results would not only have a positive epistemic effect, because it could be the basis for unbiased meta-analyses. It would also allow researchers to receive credit for relevant work they have done - that just usually would not have been published.

---

[151] http://math.rejecta.org/about-rejecta-mathematica [date of access: 26.01.2010].

[152] E.g. the Journal of Negative Results, http://www.jnr-eeb.org/index.php/jnr , NOGO, the Journal of Negative Observations in Genetic Oncology [date of access: 26.01.2010].

Such a policy would also have positive effects on the distribution of research efforts within scientific communities. If all results - irrespective of whether they were significant of not - were known, unnecessary duplication of work could be avoided. As long as non-significant results are less likely to be published, money and time may be wasted by researchers conducting studies that have repeatedly been shown to lead to non-significant results. The point I hope to have made clear is that publication practices and the space we devote and the value we ascribe to studies critically assessing others' work is crucial for science and epistemic practices more generally. The Web enables us to devote more space to various forms of critique. The tools are thus not the problem. The crucial task will be to change the differential valuation of original research and critical assessment and responses. This process can only be tackled if evaluation procedures in academia are changed accordingly. New models to evaluate scientists may need to be developed to account for and foster this change in perception: reviews and critical responses need to be sanitized and valorized. New methods may also be needed to account for and encourage more granular forms of critical practice: ratings & rankings[153], brief comments on papers should also be considered valuable contributions to science as an epistemic endeavor. Since scientists' careers depend on their evaluations, it is only if such engagement in mutual criticism receives attention and is given credit that a change in attitude may occur. On platforms, such as Slashdot.org, reviewers, i.e. people engaged in critical practice, are evaluated and receive more "Karma", for writing good comments. Positive Karma means that a user has posted more good comments than bad ones and this evaluation gives him benefits, i.e. he is only eligible to moderate with good *Karma.*[154] These comments are critical responses *and* they provide new meta-information about the quality of content at the same time. Users of Slashdot.org have clear incentives for critically assessing other people's initial postings or their comments – because their reputation, their Karma, improves and they are eligible to advance a step and become moderators, even if only temporarily. To my mind, science, and scientific publishing in particular, can only learn from such innovative practices on the Web.

[153] I conceive ratings and rankings as instances of ES: Aggregation. Hence they are depicted in more detail in the next chapter.
[154] http://slashdot.org/faq/com-mod.shtml [date of access: 13.11.2009]

Uptake of Criticism

That criticism is futile if it is not heard and does not have any effect should be obvious. Thus, for criticism to be effective, it is crucial that valid criticism is actually accounted for and leads to changes in belief. In scientific publishing, the mode of letting authors reply to critical reviews is one example where uptake is fostered. Given that on the Web not only space, but also speed ceased to be a problem, we can imagine much faster rounds of original contributions, criticism and uptake. Indeed, we do not need to imagine this - examples of almost immediate critique and responses are numerous on the Web. To the same extent to which real-time debates have long been only possible at conferences, immediate responses, replies of the authors and comments on comments are now feasible in a multitude of different ways on the Web.

However, uptake is not primarily a matter of time, but of attitude. Whether or not criticism will have an effect depends on whether I am willing to accept criticism as valid and change my beliefs and convictions accordingly. The likelihood with which I change my beliefs – or rather reveal that I have changed my beliefs - will also depend on the attitude towards change in my scientific community. There may be very well individual differences between people with respect to their responsiveness to criticism. Yet these differences may even be furthered by disciplinary standards. If consistency over time is valued higher than responsiveness to criticism, people will probably rather stick to their theories and beliefs as long as possible. If one is accused of weak-mindedness instead of being applauded for one's ability to learn and advance, uptake of criticism will be hindered. Thus, the extent to which Longino's second norm will be fulfilled in different scientific communities will depend on their attitude towards criticism, receptiveness and change – and this attitude may be more difficult to change. Thus, while the Web will be helpful in providing space for criticism, the crucial aspect of uptake will rather require shared norms of responsiveness and these cannot simply be technically implemented but need negotiations and rethinking in academia.

Public Standards of Evaluation

The third norm proposed by Longino concerns shared standards, which are temporarily binding but may change over time due to continuous scrutiny and revision. That means that these standards are developed and changed in the course of scientific practices. To use Karen Barad's terminology, they are *enacted* (Barad 2003). Although Longino does not use this terminology, I consider her account of epistemic standards to be *performative.* The performativity of these standards, the fact that they are enforced and changed through practice however, does not diminish their temporal relevance and binding force.

Clearly, without standards it is impossible to distinguish valid criticism from mere assault or nonsense. These standards mark the reference points for the evaluation of knowledge claims and the validity of criticism. Different communities may have different standards. In interdisciplinary teams that start working together for the first time, *shared standards* may still have to be developed, because the standards between disciplines differ. This need for shared standards to be developed is one of the reasons why interdisciplinary work needs a lot of time and shared space as well as initial respect and trust to work out well. For the common task they have to solve, new standards have to be developed. These norms may be explicitly discussed or implicitly enacted and performed within interactions and practices. Examples of these explicit and implicit processes have been outlined for Wikipedia as well as the Project LiquidPub before. Many norms and standards do never get explicitly declared, but rather emerge through shared practices. Despite this performative and often implicit nature of standards and norms, at times they need to be made explicit and subject to scrutiny, to be adapted and changed. This is where interdisciplinarity, or diversity more broadly conceived, looms large. Prevailing standards and the problems they might entail almost only come to the fore when different standards or norms collide. Only in encountering different norms and standards do we become aware of our own – and can make them subject to scrutiny and change.

Tempered Equality of Intellectual Authority

The final requirement concerns tempered equality of intellectual authority. While the inclusion of multiple voices and the fostering of epistemic diversity within science is a crucial goal of Longino’s social epistemology, the word tempered indicates that for epistemic endeavors intellectual authority is not uniformly distributed. In Longino’s initial declaration of the four social norms, the word “tempered” was not yet added to temper the request for equality (Longino 1990). Only later on, the word “tempered” was introduced (Longino 2002c). This seemingly innocuous change has high significance. I read this modification to denote that, despite all requests for open access and equal chances for participation in epistemic endeavors, expertise still matters and should be used to weight different epistemic agents. It is a request to recognize ability and to make epistemic use of adequate indicators of epistemic trustworthiness. This point of view is also characteristic of Scott E. Page's argument concerning the superiority of cognitive diversity (Page 2007). In computer simulations as well as in numerous empirical cases, he shows that diversity trumps individual ability over and over again. Nonetheless and despite various contradictory claims, Page convincingly argues that ability and diversity complement each other. While ability is a characteristic of an individual, diversity is a characteristic of a group and trading one for the other is like comparing apples with fruit baskets (Page 2007: xxix).

If we take a look at the Web, we can see that more people than ever before have the chance to take part in collaborative epistemic endeavors. Wikipedia, the open source movement, Slashdot.org are only some well-known examples of probably millions of epistemic communities that have formed on the Web. Of course the chances for people to take part in epistemic communities are not equal for everyone. Just looking at the digital divide between countries, between different regions within a country, between different income groups in a region should make us aware that there still are crucial differences in access that need to be adjusted. The gender ratio in different communities, such as the open source community would be another example in which differences in access may be caused less by technical access to the Internet itself than other barriers to participation. Nonetheless, joining epistemic communities, getting into intellectual exchange with others and collaborating on epistemic tasks has clearly become easier due to Web and information and communication technologies more generally.

Now that access is improved, the question of *how* to temper equality has to be tackled as well. Equality is tempered constantly by various means on the Web – several of which are outlined in the next chapter on ES:Aggregation. Indeed, I would argue that despite frequent allusion to epistemic democracy on the Web, most systems are meritocratic, if not aristocratic – or sometimes even bluntly dictatorial. As noted before, the Wikipedia predecessor Nupedia relied on academic credentials to select contributors – and failed with this strategy. Its successor Wikipedia then opened the door and allowed everyone who had access to the Internet to participate. In this case, the intellectual authority has not been tempered a priori, but quality control was introduced by different social and technical mechanisms. The ease of revision as well as the history pages are technical features that enable easy revision, the introduction of administrators with rights to block users as well as the Neutral Point of View Policy are more social mechanisms to ensure quality control.

Philpapers.org, a depository of philosophical articles from a variety of sources asks the users whether they have a PhD in philosophy to distinguish between professional philosophers and others. While the user's access to the articles depends on the personal or institutional access to the articles, the possibility of posting content depends on this difference between professional and non-professional users. Thus, the postings are tempered by the academic credentials of a user, while his access is not restricted by Philpapers.org, but only by the copyrights of the paper and the access rights of the user. Slashdot.org would be an example of a website that works quite differently in the selection of content. All users are permitted to post comments on this platform and other readers can rate these comments. These ratings are aggregated into collective judgments about the quality of these comments and those can then used by the other users to filter comments (cf. for instance Benkler 2002). As such Slashdot.org and other similar applications combine human judgment with various algorithms to evaluate content and by this "temper" the intellectual authority of the content providers. [155]

What should be obvious by now is that Helen Longino's norms can well serve as guidance for the development of socio-technical epistemic systems. In the process of design, it is possible to ask for each feature to what extent a solution supports or hinders the emergence of transformative criticism. How exactly and by which mechanism each norm has to be approached may have to be decided case by case. Yet that one should try to operationalize these norms for epistemic systems and interaction design is not only desirable but also feasible with the new information and technologies that are at our command.

Longino's approach and her four norms for transformative criticism in particular clearly demonstrate the potential fertility of social epistemology for the evaluation and the design of socio-technical epistemic systems on the Web. Her emphasis on actual epistemic practices in general and her request for local and situated modes of tempering intellectual authority makes her approach well suited to inform interaction design of socio-technical systems for epistemic purposes. Epistemology can be of use for the

[155] I return to Slashdot.org and various aggregational means to temper authority in the next chapter.

design of such systems by epistemically assessing the different features of the system. However, such collaboration does not have to stop at the stage of analysis and evaluation. A fruitful collaboration between philosophers and system designers can also lead to the development of entirely new approaches. However, the uptake of criticism and change due to new experiences should go into both directions. Not only can epistemology inform systems design – taking a closer look at different socio-epistemic mechanisms embedded in these socio-technical systems can clearly also deliver inspiration for new social-epistemic accounts of how to rationally attribute intellectual authority. I therefore conclude that applying Longino's approach is only one example for the potential fertility of bridging the gap between epistemology and systems design.

9.5 Conclusions on ES:Integration

What conclusions can be drawn from my outline of ES:Integration as a way of characterizing socio-technical epistemic systems by their mechanisms of closure? I have started with the most obvious observation, namely that systems employing this type of epistemic sociality are not only extremely prevalent, but also that they are epistemically beneficial for different reasons and in a variety of different respects and contexts. Information and Communication Technologies play a crucial role for ES^I in their ability to enable socio-epistemic processes of often enormous scale and range.

Referring to theories from social epistemology and beyond, I have outlined some key characteristics of ES^I and bundled my analysis along different topics: I have emphasized specificities of ES^I with respect to the structure of epistemic tasks and the diversity of human epistemic agents. I have further addressed the relationship between trust, trustworthiness, reputation, authority and power with socio-technical epistemic systems employing ES^I.

From the three types of epistemic sociality distinguished in this thesis, ES^I is the most common, the broadest and the most internally diversified socio-epistemic mechanism: it comes in different sizes and forms and serves various purposes. Nonetheless, it differs from the other two types which are outlined the next two chapters in various ways: they do not only function differently, they also depend on different prerequisites, they have different strengths and weaknesses, they are suited best for different epistemic tasks.

A crucial *strength* of systems employing ES^I and a major reason for its ubiquity is that it *saves time*. If a task can be split up into subtasks to be fulfilled by several people simultaneously, the overall result can be expected sooner. It is as simple as that: if two people each clean up one half of the apartment, the work will be finished earlier. Cleaning a house is an easy separable task, one person may do kitchen and the living room, the other the bath room and the bed room. Moreover, there is no problem of integration – if each task is fulfilled, the overall goal – a clean apartment – is fulfilled. Unfortunately, the tasks of division and integration are not always that easy and straightforward. Hence, system designers have to make sure that the gains in efficiency by distributing tasks over multiple epistemic agents is not annihilated by time and costs of division and integration.

Secondly, certain tasks require a diversity of skills and expertise. If there is no one to be found that possesses all these skills, the second-best solution would be to employ two people, who each possess part of the skills needed to fulfill the task. However, there are reasons to believe that employing two people with different skills may not only be the second-best, but actually the best solution to a problem. As Scott Page has shown, cognitive diversity repeatedly trumps individual ability in problem solving (Page 2007). The employment of diverse groups of people appears to be superior than searching for the one person who possesses all the skills needed for a task. This logic becomes particularly convincing if one considers that while a multi-talented genius may be well suited for the problem x, the problem she was hired for, she might be utterly unsuited for the next occurring problem y. Indeed her excellence in skills A & B might prevent her from recognizing that skill C might be needed to solve the problem y. In a cognitively diverse group of people by contrast, there may be different people who either excel in skill A, B or C – and the choice for the right strategy is more likely. Hence in that case – and in most cases – a group of people with diverse cognitive abilities and preferences, i.e. with different *perspectives, interpretations, heuristics* and *predictive models,* are more likely to find solutions for problems than individual experts (Page 2007: 7ff).

ES^I is particularly well suited for *problem solving*, as opposed to making predictions, where *ES:Aggregation* excels. Moreover, it is best at projects that *require different skills and expertise* to solve problems and that take *considerable effort*. Concerning the three central epistemic tasks, content production, evaluation and dissemination, ES^I can take place and be invaluable at all stages. In contrast to the other two types, it seems to be utile for *content production*. Many people can produce a lot more content than a single person and if this content gets combined intelligently, the epistemic benefits can be immense. ES^I clearly can also be extremely valuable for evaluation and distribution as has been shown by various examples in this chapter, but it is in content production that this type of epistemic sociality excels.

Several *prerequisites* for ES^I have also been outlined before. First of all, as all forms of epistemic sociality, it depends on trust. However, trust should not be blind and cases of scientific fraud are examples where too much trust was put into collaborators while scrutiny would have been needed. Yet without some basic trust in your collaborators, ES^I makes no sense. If one were to cross-check every result of one's collaborator – given that one possesses the skills to do so – the time savings of distributed labor would be annihilated. In fact, the working hours would double because each task would be done twice. Moreover, less measureable danger of open distrust and its detrimental effects on the working climate should be obvious and have been documented by various scholars (Shapin 1994, Kusch 2002: 75, Baier 1986).

Further, despite frequent claims to the contrary, ES^I depends on *expertise,* even if the employment of *statistical, automated control mechanisms* may diminish the risk and effect of non-experts in certain systems, such as the ones to be portrayed in the next chapter. This expertise does not have to correspond to academic credentials. As is shown exceptionally well in the case of Wikipedia, a sufficiently large number of uncertified volunteers can provide astonishing results. Nonetheless, the fact that Wikipedia is created and maintained by volunteers does not imply that they have no expertise. This expertise may be very specific and targeted, i.e. they may have a lot of knowledge about the motor of the Citroën 2CV, about the geological properties of a certain extinct volcano in the German countryside, about a certain poem by T.S. Eliot, etc. Thus, they may not need to be *certified* experts on cars, geology or English literature. Smartly combining the very specific pieces of knowledge of many individuals, coupling it with efficient technical or social quality control mechanisms seems to be more efficient for many epistemic tasks than searching for more general experts on wider fields of knowledge. Distributed expertise, abilities of varying form and degree are essential for systems successfully employing ES^I.

Obviously, ability and expertise matter. But as Page (Page 2007) asserts – so does diversity. ES^I depends on diversity. Heterogeneous groups outperform homogenous groups and individuals. Different people perceive problems from different *perspectives,* they have different *interpretations,* use different *heuristics* and employ different *predictive models.* By combining them, the space for problem solutions is increased, groupthink and incubation effects are less likely and the chances that a diverse group will come up with a solution for a given problem are higher than for individual experts or homogeneous groups (Page 2007).

That cognitive diversity is epistemically beneficial has already been defended by Mill (Mill 1859) and various philosophers of science (cf. Longino 2002c). Using Mill's arguments as a point of departure, it has been in particular Helen Longino (Longino 2002c) who has developed a social epistemology that not only stresses the necessity of cognitive diversity, but also offers normative-prescriptive guidelines on how to ensure its proper functioning.

Helen Longino's four social norms for social knowledge can serve as guidelines for structuring interaction and communication for epistemic purposes. They can be used to design platforms for interaction that enable transformative criticism and make sure that the epistemic benefits of diversity can be tapped to their full potential. Clearly, neither diverse groups in which only some members are heard and taken serious do fulfill these criteria nor groups in which everyone can speak, but no one listens. One of Longino's norms concerns the necessity of *shared standards*. This aspect corresponds to Page's notion of "fundamental preference" (Page 2007: 11). People need some basic agreement on a goal and the means of evaluating its achievement, i.e. some quality indicators and some shared vocabulary. To develop such shared standards, to be able to speak to each other and understand the other, *time, shared space* and *trust* are needed. The extent to which these prerequisites are met in different projects often decides upon the success or failure of systems depending on integrative mechanisms of closure, such as interdisciplinary projects in science.

I conclude this chapter on ES^I with a summary of the main criteria that should be fulfilled to ensure the proper functioning of systems employing ES^I. Hence, these criteria are criteria for systems analysis as much as recommendations for systems design. These recommendations are the direct result from my analyses in this chapter.

However, one has to keep in mind that there are many different instantiations of ES[I] for various purposes and of different scale, with different parameters for different epistemic tasks. Therefore, my recommendations should rather be considered as general guidelines, while their concretization needs to be specified for each case.

Setting the Parameters for ES:Integration	
Tasks: Modularity, Granularity & Integration	• Tasks should be modular and granular: The larger a projects gets, the more essential it is that the tasks are easily divisible. Varying size and content of tasks may be conducive to the attractiveness to participate and the more flexible a system is to cater to people's needs and wishes, the more likely they will contribute – and keep contributing over time. • Ease of integration is crucial. Various combinations of different integration mechanisms are possible and should be combined. Indeed, aggregative mechanisms depicted in the next chapter could be used as amendments.
Cognitive Diversity, Shared Standards & Critical Interaction	• Ability matters. Hence, making differences between epistemic agents depending on their abilities, weighting their contributions may well be epistemically beneficial. • Yet diversity matters as well. Therefore, systems should be as open as possible to enable the constitution of heterogeneous communities of cognitively diverse people who work together on epistemic tasks. • Despite their differences – or rather because of them, agents must be willing and able to communicate with one another. These interactions need time and space as well as an atmosphere of openness and receptiveness to difference, all of which need to be provided and fostered. • Despite and because of these differences, some shared goals, norms and standards for evaluation are needed. Especially the goals may be explicitly debated and agreed upon. However, many norms and standards evolve rather implicitly through practices. Nonetheless, they need to be made subject to scrutiny and this is

	where diversity looms large again. When agents with different implicit norms and standards interact, they become aware of these implicit norms and standards because they are at odds with the others' norms and standards. It is through such a clash of norms, that the norms themselves become visible and addressable. • Critical discursive interaction is needed, Longino's four social norms for social knowledge can serve as guidelines.
Trust, Authority and Mutual Effective Criticism	• Trust is indispensable for any form of epistemic sociality. In ES^I it is crucial at all instances and phases within socio-epistemic systems and processes. • Yet trust should not be blind. Loci of trust, the criteria of trustworthiness and the differential attribution of authority have to be possible subjects of scrutiny as well. • In order to be responsible knowers we have to make sure that we place trust correctly. Especially for trust in non-human epistemic agents and for forms of procedural trust transparency is crucial. • With respect to human epistemic agents we have to watch out for signs of incompetence and dishonesty. • Yet we also must be aware of the dangers of epistemic injustices that come with the differential attribution of authority.The α-error and the β-error of attributing epistemic authority.
Motivation	• Especially for voluntary projects: Choose noble goals. • Especially for voluntary projects: Make sure there is no unilateral appropriation. • Especially for voluntary projects: Ensure transparency of processes. • Make credit more modular if tasks are more modular. That is, enable the attribution of credit for tasks of different size. • Make sure that Recommendations 1-16 are fulfilled

Table 5: Setting the Parameters for ES:Integration

10 Epistemic Sociality 2: Closing Socio-Epistemic Processes Through Aggregation

10.1 Epistemic Sociality: Aggregation (ES^A)

One of the probably most well-known affirmations of the epistemic relevance of aggregational mechanisms of closure has been provided outside of social epistemology: it is James Surowiecki's book on the wisdom of the crowds (Surowiecki 2004). I have introduced his work in Chapter 8. Since some of the examples he uses to elucidate the wisdom of the crowds are indicative not only of central *features*, but also of central *differences* between various modes of aggregation, I use two of them to introduce my second type of epistemic sociality: ES:Aggregation (ES^A). These two examples offered by Surowiecki to elucidate the epistemic success of aggregation are Galton's fair visit and the story about the lost submarine.

As a reminder: Francis Galton observed the weight-judging competition when visiting a country fair. In this competition people are placing bets on the weight of an ox and the one who comes closest wins a prize. For rather political reasons, namely the question of whether people should be allowed to vote in politics, Galton was interested in the abilities of the *average person*. Hence, he used the competition to assess the average person's ability to judge oxen. This average ability to judge the weight of an ox could then be used as an indicator of the average cognitive capabilities of people more generally. To assess this average ability he collected about 800 bets and calculated the average of all the bets. Here the wonder happened: the averaged guess, the collective judgment of the crowd was 1,197 pounds – and the real weight of the ox was 1,198 pounds. Surowiecki cites Galton stating that "[t]he result seems more creditable to the trustworthiness of a democratic judgment than might have been expected" (Surowiecki 2004: xiii).

In the second example, the U.S. submarine *Scorpion* disappeared somewhere in the North Atlantic. Given the last reported location, the Navy specified a circle of about 20 miles in which the submarine probably was to be found. Surowiecki reports the story as follows (Surowiecki 2004: xxi). A naval officer construed different scenarios of what had happened to the submarine. Yet instead of having people *debate* these scenarios, he asked them to *bet* on different relevant aspects: the reasons for the trouble, the speed, the steepness of descent, etc. He then used Bayes' Theorem to aggregate these bets into a collective judgment about the location of the lost submarine -and indeed, this judgment turned out to be just 220 yards away from the spot where the ship was found months later.

What is crucial about both examples is that the collective prediction differed from all the individual predictions. The collective result was not obtained by *choosing* among individuals predictions, which would be an example of ES:Selection.[156] Nor was it obtained by *integrating* the results in an ES^I-mode, e.g. by debating and combining them. It followed a very different rationale, the rationale of ES^A.

Surowiecki states that while the two problems differ profoundly in their complexity, the basic principle to solve them is the same (Surowiecki 2004: xx). The commonality between these two examples is that in both cases, judgments of independent individuals were aggregated by statistical means to calculate "common" or "collective" results. This is the key feature of aggregation as a socio-epistemic mechanism of closure, it is the core of ES^A. For ES^A to work out, for collective judgement formed via aggregation to be superior to individual judgments – even of experts, several prerequisites have to be met concerning scale, diversity, independence and the means of aggregation. That is, for ES^A to be successful, one needs a *large enough* number of *independent* and *diverse* epistemic agents in combination with adequate aggregational mechanisms. If this is the case, ES^A will trump other modes of epistemic sociality when it comes to making predictions. The rationale behind this superiority is entirely statistical. As noted before, Surowiecki describes it as a mathematical truism. "If you ask a large enough group of diverse, independent people to make a prediction or estimate a probability, and then average those estimates, the errors each of them makes in coming up with an answer will cancel themselves out. Each person's guess, you might say, has two components: Information and error. Subtract the error, and you're left with the information" (Surowiecki 2004: 10).

However, despite this basic commonality, the examples above differ to a crucial respect: While in the case of Galton, every judgment, every vote received the same weight, in the case of the sunken submarine, different factors proposed by different people were weighted differently to judge its location. These two cases of two forms of ES^A:

[156] ES:Selection is portrayed in the next chapter.

weighted versus unweighted aggregation. In the case of the ox, this mechanism was simple averaging. In the case of the sunken submarine Bayes' Theorem was employed. While simple averaging is an example of an unweighted aggregational mechanism, Bayes' Theorem is a classic example of an algorithm to weight judgments differently. This weighting, making a difference, was already portrayed in Chapter 5, where Goldman proposed Bayes' Theorem as a means to weight others' testimony by their honesty and competence. , weighted algorithms outnumber unweighted simple averaging processes by far on the Web. This ratio between unweighted and weighted mechanisms is interesting because the term *wisdom of the crowd* often seems to imply some form of epistemic democracy where everybody on the Web has the same rights and weights. Taking a look at the actual mechanisms it becomes obvious that the Web by no means is an epistemic democracy, where every person has the same epistemic rights. Hence, it is crucial to analyze how exactly this weighting, this making of *differences that matter,* happens. Further below I outline one example for different weighting schemes in the case of recommender systems, but before I portray some examples of different systems on the Web which employ ES^A.

10.2 ES:Aggregation: Rating, Ranking, Tagging, Recommending and Betting on the Web

On the Web, there are numerous examples of systems using statistical aggregation for epistemic purposes. While some of them employ unweighted procedures, e.g. simple averaging, the majority of systems makes use of more complex algorithms to weight epistemic agents and their judgments. In the following I briefly introduce four examples of how the wisdom of the crowds is tapped via statistical aggregation on the Web: rating and ranking systems, social tagging and recommender systems and decision markets.

10.2.1 Ratings and Rankings

One of the most straightforward exploitations of ES:Aggregation are *rating and ranking mechanisms*. Ranking and ratings are means to evaluate information - or items more generally - and in doing this they create *new meta-information about the quality* of these items. In arguing that a *reputation age* is about to replace an *information age*, Origgi has noted, that the Web's epistemic relevance is not so much being a "potentially infinite system of information storage, but [rather being] a giant network of ranking and rating systems" (Origgi to appear: 2).

However, not only content itself, but also the *providers of this content* can be rated and ranked. That is, ratings and ranking provide means of classifying and assessing the quality of epistemic content *and* epistemic agents. Quite often, the processes of rating content and rating agents are combined. Reputational systems weight the content which agents provide by the reputational values of the agent. The ratings people provide for content, for instance, may be weighted by the ratings they have received as agents from others. Numerous examples of rating and ranking systems are embedded in various socio-technical systems: the rating of books on Amazon.com, the karma system of Slashdot.org, rating sellers on eBay.com, rating of products on Epinions.com. Rating and ranking mechanisms are also used on social news sites, such as Digg.com and Reddit.com, where people can submit links to articles, videos, blog entries from other sources and vote for them. The votes a story receives decide upon its position on the website. Such systems therefore link to content provided elsewhere on the Web and their own relevance lies in *attributing relevance to content* only by exploiting the wisdom of the crowds via ratings.

Such ranking and ratings provide a *way of evaluating epistemic content and epistemic agents.* As such, they are an alternative to other mechanism of communal quality control, such as *peer review*. Various tools for quality assessment exist and can and are combined in different ways on the Web and elsewhere. Yet rankings and ratings on the Web also rely on a wisdom-of-the-crowd logic, by arguing that bias gets eliminated through sheer scale. There are crucial differences between peer review and rating mechanisms and for each epistemic task the pros and cons of each type as well as of possible combinations have to be assessed. For instance, ratings on Amazon.com have a very different distribution than distributions of peer review judgments on abstract submission to conferences (Casati, Marchese et al. 2009). More specifically, Amazon ratings tend to the extreme ends of the distribution (Hu, Pavlou et al. 2009). This tendency stems from the fact that people rather rate products that they particularly liked or disliked, while mediocre products obviously do not seem worth assessing. Moreover, classical peer review of conference or journal submissions and ratings on the Web often differ with respect to exhaustion. Due to different allocation processes, i.e. self-allocation versus assignment, in peer review processes for journals, proceedings or project proposals, each submission is evaluated, while on the Amazon only a certain percentage of items get evaluated. I return to these issues as well as to the pros and cons of rating and ranking mechanisms in the next chapter.

10.2.2 Social Tagging

"Just as the internet allows users to create and share their own media, it is also enabling them to organize digital material their own way, rather than relying on pre-existing formats of classifying information" (Rainie 2007). This is how the PEW-Report on *Tagging* starts. *Social tagging* is another way to tap the wisdom of the crowd on the Web. Tagging refers to adding key words to content of the Web.

There are innumerous things that can be tagged on the Web: books (Amazon.com), articles (CiteUlike.org), pictures (Flickr.com), websites (Delicious), videos (Youtube.com), blog entries (Technorati.com), short text messages (Twitter.com), music (last.fm). Tagging provides a means to classify and categorize information for later information retrieval. In contrast to top-down classifications where categories into which material can then be sorted are predefined, tagging is a bottom-up approach. The users choose the tags not from a fixed set of possibilities, but by using a phrase that they think classifies the content to their requirements.

In direct reference to the Dewey Decimal System, a library classification system, the PEW stresses that tagging is rather "tailored to individual needs and not designed to be the all-inclusive system that Melvil Dewey tried to create with his decimal-based scheme for cataloguing library materials" (Rainie 2007). Yet in most systems tagging is not a purely individual activity, but has a clear social component. Think of tags as public keywords. In epistemic social software your tags are added to the tags others have used to describe content. They are used to build up a repository of second-order information about content on the Web, often referred to as *folksonomies.* By doing this, they help organizing knowledge in a quite different way than conceived in traditional knowledge organization. As David Weinberg puts it in an interview, "[f]olksonomies reveal how the public is making sense of things, not just how expert cataloguers think we *ought* to be thinking" (Rainie 2007).

Moreover, by tagging content systems often create recommendations on how to tag content for other users by offering the "most popular"-tags for a given website, picture or whatever content you wish to tag. Whenever users want to save a website on their Delicious-account, they are offered the popular and recommended tags. Popular tags are calculated by the tags others have applied to content, while recommended tags are a combination of tags one has used to tag *other content* and tags that *other people* have used[157]. On Flickr, popularity of tags is indicated by a tag cloud, in which the font size of the word for the same content indicates the frequency by which this tag has been used.[158] This tag cloud thus visualizes the popularity of tags and enables users to search for content which was tagged by these most popular tags. Flickr, a website to store and share pictures, has another interesting feature to classify content, namely the category of *interestingness*. Here is how Flickr defines interestingness:

> "There are lots of elements that make something 'interesting' (or not) on Flickr. Where the clickthroughs are coming from; who comments on it and when; who marks it as a favorite; its tags and many more things which are constantly changing. Interestingness changes over time, as more and more fantastic content and stories are added to Flickr."[159]

Interestingness is a new category based on the assessment of content through users. It combines the number of tags, which are *content qualifiers*, with other criteria, which are rather *quality qualifiers*. Hence while the content of the tags refers to the picture's content, the number of tags together with the number of clickthroughs, the number of comments and times it has been marked as favorite offers yet another way of classifying content, namely by its interestingness.

Another aspect of Flickr's calculation of interestingness is worth mentioning. Its description quoted above implies that not only the *number* of tags, clickthroughs and markings as favorite matter, but also *who* performs this action: Interestingness is a feature of content that depends also on "[*w*]*here* the clickthroughs are coming from; *who* comments on it and when; *who* marks it as a favorite".[160] Thus, although the concrete algorithm of how interestingness is calculated is not outlined on Flickr, its description seems to imply the interestingness-feature makes use of reputational mechanisms, of making differences between users that have an effect on the weight of their tags, marks and clicks.

[157] http://delicious.com/help/faq#tags [date of access 20.3.2010]
[158] http://www.flickr.com/photos/tags/ [date of access 20.3.2010]
[159] http://www.flickr.com/explore/interesting/ [date of access 20.3.2010]
[160] http://www.flickr.com/explore/interesting/, emphasis added. [date of access 20.3.2010]

10.2.3 Recommender Systems

Another example that relies crucially on aggregation of data from a multitude of different people are *recommender systems (RS)* based on collaborative filtering. Recommender systems suggest items to a user of a system that he or she may be interested in. A classical example are recommendations you get on Amazon.com. Based on your previous purchases, the items that are on your wish list and those you may have rated, Amazon.com sends you emails with books or other products you may be interested in. While one may well argue that the only reason why Amazon.com is using such a system is because they want to sell more books, for the user there may be an *epistemic advantage* of recommender systems. Recommender systems are an automated form of giving advice that is based on different criteria, amongst others the aggregation of quality judgments obtained from numerous agents.

As noted before, there is a clear epistemic benefit in finding information one is looking for fast and reliably. That is the logic behind the epistemic utility of good search engines. However, there may be an even bigger epistemic benefit, if a system can satisfy an epistemic need that one was not even aware of. If a recommender system draws my attention to something that is relevant for my interests, but that I was not even looking for or considered to be relevant before. Collaborative filtering processes and different statistical procedures may be particularly useful in providing such unexpected, relevant information. Therefore, RSs can be considered socio-technical systems that serve epistemic purposes. The *social* part of this *socio*-epistemic mechanism lies in the fact that for collaborative filtering to make recommendations for you, your data are compared to that of other users in order to find out which patterns resemble yours. The *epistemic* part resides in the fact that they are trying to *predict* your interests – and in the possibly epistemically beneficial consequences of being able to offer something of interest one was not aware of before, but may need for your epistemic inquiry. A special type of RSs, trust-aware recommender systems exploit epistemic sociality in even another respect. Trust-aware RSs also make recommendations, but they are based on even more social information. “While traditional RSs exploit only ratings provided by users about items, Trust-aware Recommender Systems let the user express also trust statements, i.e. their subjective opinions about the usefulness of other users” (Massa and Avesani 2006). I return to trust-aware RSs in more detail in my case study below.

10.2.4 Prediction Markets

Bragues, who addresses the socio-epistemological aspects of prediction markets, defines them as "[...] venues in which individuals trade securities whose value is tied to the outcome of a future event (Bragues 2009: 93).[161] By structuring prediction markets as betting exchanges, the existing market price can be interpreted as a prediction of the probability of an event. Numerous prediction markets exist on the Web. One of them is the Hollywood Stock Exchange[162], in which actors, directors or films can be traded and bets can be placed on the next Bond girl. Another prominent example is the Iowa Electronic Markets, which describes itself as "an on-line futures market where contract payoffs are based on real-world events such as political outcomes, companies' earnings per share (EPS), and stock price returns".[163] An overview over the most important prediction markets is given by Bragues (Bragues 2009).

However, to interpret the prediction market prices as probabilities, several prerequisites have to be met, such as *liquidity*, i.e. the ability to attract enough participants (Bragues 2009). Moreover, the participants should possess *diverse opinions and cognitive strategies* (Page 2007) *and should act independently* to avoid information cascades (Coady 2006, Bikhchandani, Hirshleifer et al. 1992). These prerequisites are hence quite similar to Surowiecki's prerequisites for a "wisdom of a crowds" (Surowiecki 2004): diversity, independence, decentralization and modes of aggregation. Another factor that affects the validity of interpreting market prices as probabilities concerns the risk-aversion of the traders. According to (Wolfers and Zitzewitz 2007), "[g]reater risk-aversion leads to a bias toward more extreme prices, while lesser risk aversion leads prices to be biased toward $0.50". Nonetheless, they conclude that "[i]n most cases [...] prediction market prices aggregate beliefs very well. Thus, if traders are typically well-informed, prediction market prices will aggregate information into useful forecasts. The efficacy of these forecasts may however be undermined somewhat for prices close to $0 or $1, when the distribution of beliefs is either especially disperse or when trading volumes are somehow constrained, or motivated by an unusual degree of risk-acceptance" (Wolfers and Zitzewitz 2007).

---

[161] Bragues' (2009) analyses are depicted in Chapter 6.
[162] http://www.hsx.com/ [date of access: 21.03.2010]
[163] http://www.biz.uiowa.edu/iem/index.cfm [date of access: 21.03.2010]

In comparing the utility of this ES:A-based form of prediction with traditional expert-based predictions, Bragues concludes that while "[...] in non-empirical disciplines like literature and philosophy, it has long been customary to rely on individual geniuses to provide illumination, [...] prediction markets disclose that there is much knowledge to be gained by analyzing the fruits produced by the combined efforts of many people seeking to comprehend the same problems "(Bragues 2009: 103).

10.3 ES:Aggregation and Social Epistemology

While topics related to ES:Integration have been widely addresses in social epistemology, aggregational mechanisms have been of lesser concern. Nonetheless, there are socio-epistemological analyses on how to use aggregational mechanisms for epistemic purposes. One of the most important ones has been introduced before: Lehrer and Wagner's rational model of consensus. In Chapter 6 I have introduced Keith Lehrer's model in some detail within the section on consensus as a socio-rational mechanism. In this chapter, I want to argue that this model is a prime example of ES:Aggregation in a dual sense, because it applies aggregational mechanisms to *epistemic agents and to epistemic content*. Aggregational mechanisms are first used to assess the reputational weights of epistemic agents by letting epistemic agents assess and re-assess their peers' epistemic trustworthiness in several feedback loops. Once these reputational weights are obtained, they are used to weight the answers which epistemic agents have given on some question in dispute. These weighted judgments are then aggregated to form a collective result. Hence, Lehrer's model is an excellent example of process in which a weighted form of aggregation is applied to content and agents – a process which is highly similar to the function of reputational mechanisms on the Web.

10.3.1 Lehrer's Consensus Model as Epistemic Sociality^2:Aggregation^2

Recall the basic idea of Lehrer & Wagner's model (Lehrer and Wagner 1981): At first you let a community of peers collectively assess and consent about the intellectual authority of each peer within this community with respect to a specific topic. Each expert is asked to make a judgment concerning the perceived level of expertise her fellows have with respect to a specific question at hand. This judgment is a quantitative indicator summarizing all her information about the other's expertise and reliability concerning the issue at stake, in other words: it gives a quantitative indicator of what she considers to be the reputation of each scientist with respect to a topic. Then the average reputation values for each scientist are calculated with a specific algorithm and send back to all members. In a second round, all experts have to reassess their initial judgment taking into account the averages indicating the amount of intellectual authority that their fellows ascribe to each other.

The idea is that the less secure someone is about a peer, the more likely it will be that he corrects his judgment towards the group average. If he is very sure about the competencies – or lack thereof – of a peer, the average will not influence his decision too much, because he considers his own information to be more important. If these processes were repeated endlessly, the values would finally stabilize and a consensual weight for each member of a community would be achieved. This consensual weight would be a quantitative indicator of the intellectual authority that the community as a whole ascribes to each member; it is a quantitative indicator of someone's *reputation*[164]. Once this value is there, it can then be applied to weigh the judgment of each member on a factual matter. In other words: after having asked the community members to assess each other repeatedly, one finally asks them about their opinion on the question of interest. Then these judgments are aggregated weighed by the reputation score of those who made the judgments.

Although this process is similar to Delphi methodology (Linestone and Turoff 2002), there is one crucial difference. In Lehrer's model, topic matters are not assessed right away. Instead, he develops a *reputational map of an epistemic community* first, which is then used to weigh the answers to topic matters. Aggregation serves the epistemic purpose of making decision making more rational and is done at two stages: first an aggregational mechanism is used for assessing the reputation values of the experts (or the consensual weights in Lehrer's terminology) and then the judgments on the topic of concern are also being aggregated. After having used ES:Aggregation to assess the

[164] While I adopt the term *reputation* here, other authors have offered different terms to label Lehrer's social information. Goldman (2003: 76) for instance, uses the terms *respect* and *trust.* Clearly, all these terms have something in common, namely that they are evaluative information about other people. However, trust, reputation, respect, or social information are not used consistently within or between different fields of research. Thus, a clearer terminology of the differences between these notions would be a first step in improving the discussions about these topics.

reputation of each epistemic agent, one uses ES:Aggregation again to aggregate their judgments on a topic matter to form a collective result.

This model of rational consensus thus is a socio-rational mechanism in which both sociality and aggregation are doubly represented. It is a case of "ES^2:Aggregation^2". Not only does this model employ a *social mechanism* to achieve consensus on a factual matter. To decide on the factual matter this model makes use of *social information* defined as the opinions experts have of each other. Moreover, Lehrer's model is an example of the epistemic use of aggregation, also in a dual sense. Not only are the community members' opinions of each other aggregated. In the end, their opinions on a topic matter are also being aggregated - weighed by the reputation they have been ascribed by their peers. Hence, Lehrer's model can serve as a prime example of ES:Aggregation: it uses *social information* and *aggregational mechanisms for epistemic purposes*. An interesting aspect of Lehrer's model concerns the fact that it provides an example in which *experts are being aggregated*. Thus, it delivers a model in which expertise is not in contrast with aggregational mechanisms, quite to the contrary: expertise itself is assessed by aggregating experts' judgments.

One central insight to be obtained from Lehrer's analyses is that reputation, information about other people's quality as knowers, about their competence and honesty, is of epistemic value. It is not only frequently *being used* to assess the epistemic claims of epistemic agents, it can also be *useful*. Of course, not all epistemic usage of reputational cues has to follow such a formal method. Mostly we use reputational cues rather implicitly. Nonetheless, reputational mechanisms are formalized and embedded in the majority of Web2.0 applications. Ratings, rankings and other reputational tools are used in a variety of different ways on the Web, and Lehrer & Wagner's model delivers one early formalization of how to use reputation for epistemic tasks (Lehrer and Wagner 1981).

Nonetheless, there are certain problems with Lehrer's model, some of which have been outlined in Chapter 6. Indeed, Alvin Goldman, who not only advocates for the use of reputation to weight others testimony, but also proposes a statistical mechanism for this consideration, has outlined two major points of critique from a socio-epistemological perspective.[165] The first question concerns the adequacy of Lehrer's model as a social

[165] Confer Part 2 of this thesis for a depiction of Goldman's Bayesian account of testimony.

epistemology; the second concerns the origin of the initial weights. Concerning the first issue Goldman argues that as an example of pure proceduralism, Lehrer's model does not provide an adequate basis for a comprehensive social epistemology and states: "Many important social activities performed by scientific and other communities cannot be addressed within this framework, such as coordinated evidence gathering, proposing and rebutting arguments, allocating speech opportunities, and creating incentives for research and investigation. In short, the vast majority of social practices that belong under social epistemology fall outside the purview of this approach" (Goldman 2003: 77). This critique refers to the fact that neither Lehrer's model - nor ES:Aggregation in generally – exhausts the realm of epistemic sociality. This is exactly the reason why I am proposing a tripartite notion of epistemic sociality to form the basis of any comprehensive social epistemology. In such a view it is possible to assess the merits of different theories for explaining different socio-epistemic mechanisms even if they do not deliver full-fledged accounts of epistemic sociality. Thus, Lehrer's model can be put in place and be evaluated only with respect to its utility of exploiting ES^A. After all, only few theories, if any, aim at accounting for the sociality of knowledge in all possible forms. Indeed, one crucial argument of this thesis is that most social epistemologists focus only on certain socio-epistemic practices while neglecting others. My tripartite notion of epistemic sociality delivers a framework for classifying and assessing the respective merits of these theories for different epistemic tasks. The fact, that neither Lehrer's consensus model nor Goldman's own account exhaust the realm of epistemic sociality does not imply that they are not epistemically useful.

Goldman's second point of critique is more crucial and refers to a blind spot of Lehrer's model that has been mentioned before: the origin of the initial reputation values. Lehrer's model "does not address the general question of how to assess the expertise or competence of others, but simply adopts subjectively chosen assessments as givens" (Goldman 2003: 77)). As any mathematical procedure it depends on the quality that is fed into it. Yet, while Goldman rightly points out the neglected origin of the initial weights in Lehrer's model, Goldman's own account of the origin of initial values needed for Bayesian inference in his model of testimony is not too convincing either (Goldman 2003: 123f). While he notes that hearers may use track records to assess speaker's competence and honesty, he also acknowledges the limits of track records when he writes: "The significance of track records, however, should not be exaggerated.

Although Jones never lied before, perhaps he never had as strong an incentive to lie as he does now" (Goldman 2003: 124).

Nonetheless, with his critique Goldman points to an important issue: taking a close look at the initial data which are fed into reputational systems and at the algorithms that are employed to process these data should therefore be a crucial task for any socio-epistemological analysis of socio-technical epistemic systems using reputation tools. Unfortunately, these algorithms are not visible to the average user or even not disclosed by the system developers at all for better (e.g. to avoid system attacks) or worse reasons. However, considering the substantial effect that such reputational mechanisms have on the information we receive and the weight that is attributed to different users, that is the differential effects they have on *epistemic content and epistemic agents*, these mechanisms need to be made transparent and subject to close scrutiny.

10.4 Key Topics

10.4.1 ES^A & Epistemic Tasks

From the examples outlined in this chapter we can see that aggregational mechanisms are employed for numerous tasks on the Web. More specifically, they are employed for the following epistemic tasks:

a) to *select and filter content* by aggregating *quality judgments* about epistemic content and agents in the case of ratings and rankings as well as recommender systems;
b) to *weight* epistemic agents and their judgments by aggregation quality judgments about epistemic agents and the content they provide;
c) to *save and retrieve information* and create bottom-up classification systems (folksonomies) by aggregating content qualifiers in the case of tags;
d) to make *predictions* by aggregating judgments either directly on events or on topics relevant for this event in the case of prediction markets.

Given these specificities, ES^A is suited for different epistemic tasks than ES^I. Its capacity to create new knowledge is restricted to very specific forms of knowledge, such as predictions, or second-order knowledge, such as classification systems. Closure through aggregation does not result in new epistemic products that depend on the creation and integration of diverse epistemic content, as was the case for ES^I. Its strength lies rather in assessing content with respect to various possible criteria such as quality, adequacy, type, topic, etc. For each of these tasks, the merits of ES^A have to be compared with those of other mechanisms. One reasonable rationale for this comparison could be Goldman's five standards of epistemic appraisal: reliability, power, fecundity, speed and efficiency (Goldman 1992: 195).

With respect to *classificatory work*, folksonomies could be compared to top-down-classification systems regarding their reliability, power, fecundity, speed and efficiency. Indeed, there is considerable work on folksonomies in the library and information sciences and the field of knowledge organization.[166] While speed, efficiency and fecundity may clearly be advantages of folksonomies, the issues of reliability and power are harder to tackle and subject to much debate (e.g. Shirky 2005). Different authors have therefore worked on intermediary solutions that try to combine tagging with top-down classification systems (e.g. Hidderley and Rafferty 2006).

Concerning the tasks of *filtering and selecting content*, ES^A-based processes could be compared with different forms of peer review. Indeed, analyses on peer review often use similar criteria to assess the merits and flaws of different peer review processes. (Casati, Marchese et al. 2009) analyze double-blind peer review processes with respect to their ability to select high quality papers in a process that is fair and efficient. Hence, speed and efficiency play a crucial role while reliability is even more central. Indeed, it is the reliability of peer review that is currently being questioned by different researchers. Having analyzed blind peer reviews of conference submissions, Casati, Marchese et al. (Casati, Marchese et al. 2009) conclude, for instance, that the review process has performed very poorly if future citations of the papers are used as a criterion for their quality. According to them a random selection would have had "the same quality in approximating the citation-based ranking" (Casati, Marchese et al. 2009). Moreover, their research shows that the selection process is crucially influenced by *rating biases* of the reviewers, i.e. the difference between raters who consistently give higher or lower marks than their colleagues when assessing proposals. I return to this topic in the next chapter in more detail. It is in particular because of such critical analyses of the merits of

[166] Confer for instance, the proceedings of the numerous international and national conferences organized by the International Society for Knowledge Organization. Further information on http://www.isko.org/ [date of access: 23.01.2010].

peer review that alternative methods of selecting for quality in general and methods that employ ES^A in particular are currently being assessed. Some work on comparing ES^A-style selection mechanisms in science with traditional peer review is currently being conducted within the Project LiquidPub[167], but the results are not yet clear. To my mind, the future of quality assessment in science will lie in combined methods rather than in a replacement of traditional peer review with rating or ranking mechanisms. This is not only the case because peer review serves different quality control functions beyond selection (such as detection of fraud, making suggestions for improvement or future research). It is also not yet clear how it is possible to distinguish quality assessment from mere popularity – and how this difference should be conceptualized in the first place. Hence, there are many issues that have to be addressed to assess the comparative merits of ratings and ranking mechanisms and of ES^A for selecting content for its quality. What can be said at the very least is that aggregational mechanisms provide a *fast* way of assessing content.

Finally, ES^A not only provides a fast way of assessing content, it also provides fast mechanisms of *taking decisions* – and *making predictions.* As compared to possibly protracted deliberative processes, simple betting or voting procedures provide faster ways of reaching results. With respect to making prediction, aggregational mechanisms are not only faster, but also more accurate than other mechanisms as has been shown for prediction markets. Prediction markets, by interpreting prices as probabilities of future events frequently outperform expert judgments. Hence, there seems to be a clear epistemic benefit of employing ES^A for epistemic tasks that involve simple predictions (e.g. the ox weight) as well as more complex predictions (e.g. the lost submarine (Surowiecki 2004)), which may include the weighing of different factors. This superiority of ES^A, however, is bound to certain prerequisites andin contrast to expert judgments, these prerequisites are not characteristics of the individual epistemic agents (e.g. their competence), but characteristics of the sum of all epistemic agents: the number, diversity and independence of epistemic agents.

[167] http://project.liquidpub.org/ [date of access 23.01.2010]

10.4.2 ES^A: Numbers, Diversity, Independence & A Statistical Style of Reasoning

The proper functioning of ES^A depends on several prerequisite. Thinking back to the rationale of Surowiecki's wisdom of the crowd (Surowiecki 2004), these prerequisites can be summarized as follows: You need *numerous, diversely* skilled and informed epistemic agents, who act *independently* before their epistemic content (e.g. quality judgments, content classifiers, etc.) gets aggregated with an *appropriate aggregation mechanism*. If these prerequisites are not met, different types of problems occur.

If the epistemic agents involved in ES^A are not acting *independently*, information cascades can be the consequence.[168] The problem with information cascades is that error can be spread if people rely on others' information more than they do on their own information. The basic rationale behind aggregation as an epistemic mechanism is that, if a large enough number of people is asked the same question, their error may be cancelled out and one is left with the information - to use information theoretical terminology. However, if people do not rely on their own information, but on the *error* provided by others, this rationale is violated. In laboratory situations, people may be asked to place their bets simultaneously to avoid information cascades. However, such simultaneity may be difficult to achieve outside of laboratory settings and other means may have to be employed to avoid information cascades (Coady 2006).

The second problem concerns groups in which epistemic agents may act independently, but lack the necessary *diversity*. Similar people, who use the same information or the same heuristics and cognitive strategies, may induce error into the averaging process, because they are all *biased in the same way*. In statistical terms this means that one crucial prerequisite for classical test theory is violated, namely that error is random. If error is not random, the process of cancelling it out via statistical means does not function anymore. It has to be noted that the notion of bias that is used here is an entirely *statistical* notion of bias. Hence, it has to be distinguished from other types of bias to be found in ICT, such as the differentiation into *societal, technical of emergent bias* proposed by (Friedman and Nissenbaum 1997). This characterization of bias as statistical bias brings us to a more general issue: the relationship between ES^A and statistics. Ian Hacking has introduced the term *statistical style of reasoning* (Hacking

[168] Information cascades have been introduced in Chapter 8.

1992a) and I want to argue that a) ES^A is indicative of such a statistical style of reasoning and b) that the utilization of ES^A-type mechanisms in many socio-technical epistemic systems on the Web leads to a reinforcement and amplification of this statistical style of reasoning.

What is a *style of reasoning*? Hacking does not precisely define what it is, but only refers to its metaphysical and performative nature by stating that a style of reasoning is “an irrevocably metaphysical idea, yet styles like all else human, come into being through little local interactions“ (Hacking 1992a: 131f). Due to this lack of a definition, taking a look at the *examples* of different styles of reasoning he distinguishes as well as the *effects* he is ascribing to them is more illuminating.

In the development of his concept of styles of reasoning, Hacking refers to two predecessors: A.C. Crombie *styles of scientific thinking,* based on historical analyses of scientific practices (Crombie 1994, Crombie 1988), as well as Ludwik Fleck’s *thought styles* (‘Denkstile’), which result from his sociological analyses in medicine (Fleck 1980). Since Hacking’s style concept can be considered as a philosophical counterpart to Crombie’s historical concept, Crombie’s six styles should be mentioned. Hacking summarizes them as follows:

> “(a) The simple postulation established in the mathematical sciences.
>
> (b) Experimental exploration and measurement of more complex observable relations.
>
> (c) Hypothetical construction of analogical models.
>
> (d) Ordering of variety by comparison and taxonomy.
>
> (e) Statistical analysis of regularities of populations, and the calculus of probabilities.
>
> (f) Historical derivation of genetic development.” (Hacking 1992b: 132).

Hacking adds to this list the *laboratory style of reasoning*, which is “characterized by the building of apparatus in order to produce phenomena to which hypothetical modelling may be true or false, but using another layer of modelling, namely models of how the apparatus and instruments themselves work” (Hacking 1992b: 6). However, for the purpose of this thesis, it is the *statistical style of reasoning* is the central style of reasoning which is also the one Hacking has developed in most detail based on his analyses on the historical development of statistics (Hacking 1992a, Hacking 1975; Hacking 1990).

Yet why introduce a concept of style? What is supposed to be explained by it? Hacking and his predecessors use the concept of styles to shed some light on knowledge creation in science. To my mind, the style concept is a way to describe different epistemic practices and explain how they function. It offers a way of grasping the *relationship between epistemic practices and epistemic content* and as such it is helpful to understand the *performative* relationship between epistemic practices on the Web, the epistemic *content* that gets created by those *practices* and the epistemic *norms and standards* that evolve within and through epistemic practices. This relevance becomes obvious when taking a look at three core characteristics of styles.

First of all, a style of reasoning according to Hacking introduces „new sentences with new meanings, new truth-conditions, new objects, new classifications, and new criteria for verification" (Hacking 1992a: 142). That is, each style of reasoning, for example the statistical or the laboratory style of reasoning, creates its own objects of studies, new classes, types of propositions, laws and explanations, etc. Secondly, styles of reasoning are *self-authenticating*. They define their own criteria of validity and objectivity, their set of propositions that can be true or false. Hacking states that "each style of reasoning introduces its own criteria of proof and demonstration, and that it determines the truth conditions appropriate to the domain to which it can be applied. [...] A style of reasoning is more than a group of techniques for bringing new kinds of facts into our awareness, into our living, mental, social world. I say it creates the very criteria of truth. It is, as I like to say, self-authenticating" (Hacking 2002: 4). For the statistical style of reasoning this means for instance that measurements of accuracy, reliability and even validity are based on statistics itself (Hacking 1992a: 144).

Third, styles are also *self-stabilizing*, i.e. they develop their own techniques, by which they stabilize themselves. In that sense styles are validated through their application. This aspect is crucial for my argument that currently a statistical style of reasoning gets amplified by the sheer ubiquity of statistical mechanisms on the Web. Different styles have different techniques of stabilizing themselves and these techniques are also differently successful, but their commonality is that "they enable a self-authenticating style to persist, to endure" (Hacking 1992b: 15). Hacking has analyzed these techniques for three styles he distinguishes: the statistical, the laboratory and the mathematical styles of reasoning (Hacking 1992a, Hacking 1992c, Hacking 1995, cited from Hacking

2005).[169] The statistical style has been particularly successful in stabilizing itself and Hacking notes:

> "the statistical style is so stable that it has grown its own word that gives a hint about its most persistent techniques: 'robust'. In the case of statistics there is an almost too evident version of self-authentication (the use of probabilities to assess probabilities). But that is only part of the story, for I emphasize the material, institutional requirements for the stability of statistical reasoning. Indeed, if my accounts deserve to be pegged by any one familiar philosophical '-ism', then it is materialism."(Hacking 1992b: 15f).

I would like to add that when it comes to the Web, this materialization takes place by inscribing statistical mechanisms into socio-technical epistemic systems. It is by this materialization that the statistical style of reasoning is performed and boosted.

The statistical style of reasoning has long left the realm of science and arrived in the midst of our daily life. According to Hacking this statistical style of reasoning "has totally changed our feel of the daily world in which we live, a world in which everything is cloaked in probabilities, sex, sports, disease, politics, electrons, cosmic collision, the wave function. The triumph of probability was engineered in the nineteenth century and perfected in the twentieth. Now it is inescapable" (Hacking 2002: 3). I argue that the Web2.0 environment has given this development yet another spin: the sheer *ubiquity* of tools that employ statistical aggregation has boosted the perceived *value* of statistical aggregation. That is, a *statistical style of reasoning* is being reinforcing through its ubiquity in various fields of our daily life in general as well as through its embeddedness in many socio-technical epistemic systems in particular. Some support for this claim can also be found in debates about the end of expertise on the Web and beyond, which is either welcomed (e.g. Surowiecki 2004) or feared (e.g. Keen 2008). Numerous books have become bestsellers in the last years, which can be interpreted as promoting or assessing the merits or shortcomings of statistical reasoning (e.g. Ayres 2007, Anderson 2006, or Taleb 2007 for a different perspective).

By referring in particular to prediction markets and evidence concerning the comparative advantages of statistical mechanisms over individual expert judgments, several authors from primarily economic or technical backgrounds have argued for the replacement of expert judgments with statistical techniques (Surowiecki 2004, Ayres

[169] In contrast to Crombie (1994), Hacking does not consider these styles to chronologically displace each other. Indeed, in the case of science, he argues usually several of these styles are at work simultaneously.

2007). In the following I use Ian Ayres's bestseller "Super crunchers: why thinking-by-numbers is the new way to be smart" (Ayres 2007) as an example to show not only how a statistical style of reasoning is defended, but also how this defense is related to ICT.

Ian Ayres delivers an account of how large-scale, statistical, data-driven procedures can solve Surowiecki's cognition problems (Surowiecki 2004). The title "Super Crunchers" refers to data set analysts who process huge amounts of data for decision and prediction making: which baseball team is going to win this season? How good – and expensive - will a wine be that has just been harvested? Which teaching model is superior? What illness is a patient suffering from? The bottom line of Ayres argument is that each of these questions can better be answered by using statistical mechanism than by asking experts.

Ayres lists numerous examples in which statistical predictions based on large data sets outperform expert judgments based solely on intuition and observational experience and concludes: "For decades, social scientists have been comparing the predictive accuracies of Super Crunchers and traditional experts. In study after study, there is a strong tendency for the Super Crunchers to come out on top" (Ayres 2007: 108). The reason for this superiority lies in the fact that in contrast to algorithms, humans often falls prey to certain cognitive biases, irrespective of whether they are experts or not (Ayres 2007: 115f): we often put too much weight on unusual events, we have a tendency to cling to beliefs we have formed even if they turn out to be false, and we are overconfident in our cognitive abilities. Moreover, we are not particularly good in putting weights to different causal factors – and this is a task in which quantitative data analyses are much superior to our intuitions. Since algorithms do not have an ego, they also neither cling to mistaken beliefs not do they overrate their own abilities. The main conclusion that Ayres draws from these observations is that "we should strip experts of at least some of their decision-making authority" (Ayres 2007: 116). Instead of making decision, the role of experts should rather be elsewhere: someone has to come up with hypotheses to be statistically tested, data needs to be collected, created and interpreted. Yet when it comes to deciding and predicting, it is the algorithm that should have the last say, because "[h]umans too often wave off the machine predictions and cling to their misguided personal convictions" (Ayres 2007: 117). Ayres is explicitly arguing to shift authority from human epistemic agents to algorithms and in the next section on trust, authority and transparency I will show that this shift, this change towards *algorithm authority* (Shirky 2009) has already been executed on the Web.

According to Ayres, there are two crucial prerequisites for *Super Crunching* to emerge: statistical techniques and information and communication technologies to provide and process data. While several of the statistical techniques have existed since centuries, it is the technological change, the availability of ICT, which has enabled *Super Crunching* – and has thus boosted a statistical style of reasoning. Indeed, many of examples that Ayres uses to explain the merits of *Super Crunching* are taken from the Web.

With respect to the statistical basis, Ayres refers to two basic techniques and one theorem: *randomization* and *regression analysis* as the basic techniques for making predictions and *Bayes' Theorem* as a tool for updating predictions over time. Regression analysis produces an equation that best fits a set of given data. However, once the equation is there, it can then also be used for making predictions. One example that Ayres uses to explain how regression analysis functions is the online dating service E-harmony.[170] Based on analyzing various traits of married couples, E-harmony has created a model of twenty-nine variables with different weights that should predict the match between two singles in the E-harmony database (Ayres 2007: 22ff). In principle, this model draws conclusions about the relevance of different personality factors for the compatibility of partners based on the sample of 5000 married couples, i.e. successful examples, to make predictions for those singles still waiting for their match. That is, historical, existing data from married couples are used to make predictions for the singles. In contrast to many recommender systems, decision making and predictions in the case of this dating service does not depend on explicit, conscious preferences. Rather, the model seeks to reveal some *underlying and unknown factors* (Ayres 2007: 27). Another crucial characteristic of regression analyses are confidence intervals. Confidence intervals describe the range into which a value falls with a likelihood of 95% (or 99%). That means that regression analyses deliver for each value also an indicator of the reliability of this value. Depending on the quality of the input data, this confidence interval can be larger or smaller. The narrower a confidence interval is, the more precise the prediction is.

[170] http://www.eharmony.com/ [date of access: 19.11.2009]

Randomization is the second technique that Ayres refers to as the gold standard for data-driven decision making. Let's assume that you want to find out whether a drug is effective in treating a certain disease.[171] You may use two groups: the experimental group, which receives the treatment, and the control group, which receives a placebo. In order to prove that differential results are really due to your treatment, you have to make sure that there are no *other variables* that could explain for the difference. One possibility for this is to *match* the groups, i.e. you would have to make sure that the people in the two groups are similar with respect to all other variables that may have an effect of the outcome other than your treatment. With randomization you avoid difficult matching procedures. After all, you would have to know which ones are the relevant variables necessary for matching, and even if you knew them, this would pose additional problems for acquiring participants for your study. Thus, instead of matching, one uses randomization, which relies on the same logic that Surowiecki has proposed for his wisdom of the crowds: if you ask a large enough group of diverse, independent people, then their errors will cancel each other out (Surowiecki 2004: 10). Size reduces bias if certain prerequisites are met. The same is true for the different variables affecting the outcome of a test. If the sample size is large enough, randomization makes these *groups* identical. Instead of matching *individuals, distributions* are matched. Ayres concludes: "Since the *distribution* of both groups becomes increasingly identical as the sample size increases, then we can attribute any differences in the *average* group response to the difference in treatment (Ayres 2007: 51). So all we need for number crunching are sufficiently large sample sizes, i.e. large data sets.[172]

The final tool that Ayres emphasizes has been introduced before: the *Bayes' Theorem* as a model of how to update probabilities in the light of new evidence. Ayres argues that "the Bayes equation is the science of learning. [...] If the Super Cruncher of the future is really going to dialectically toggle back and forth between her intuitions and her statistical predictions, she's going to have to know how to update her predictions and intuitions over time as she gets new information. Bayes' equation is crucial to this updating process" (Ayres 2007: 212).

[171] For the sake of simplicity, I am only portraying a much simplified version of actual clinical test trials of course.
[172] However, it should be noted that the utility of statistical significance, i.e. the ubiquitous p.05 and p. 01, for large-scale data analyses has recently become under dispute. Confer (Sterne and Smith 2001) or http://www.hastac.org/blogs/cathy-davidson/statistical-significance-v-substantive-significance for further literature [date of access: 23.03.2010]

As stated before, these statistical techniques are not at all new, the Bayes' theorem for instance was formulated in the 18th century (Bayes 1764). It is not techniques, but *technology,* and more specifically *computer technology*, that makes the difference in two respects. First of all, the exponential increase in processing and storage capacity has lead to *more data* that can be *processed faster*. Thus, it is not due to new statistical methods, but due to the availability of data and means of processing them that *Super Crunching* has emerged. Moreover, in the Web2.0 environment there are not only applications that create data, but many applications process these data themselves and use them for a variety of different purposes. That is, ICT and computer technology more generally serve as prerequisites, amplifiers and playground for *Super Crunching*. Indeed, as has been shown throughout this chapter, many socio-technical epistemic systems employ aggregational mechanisms. These mechanisms range from simple averaging processes to numerous more complex algorithms embedded in different types of systems: the filtering, selection and evaluation of epistemic agents and content, classifications, recommendations, search results – on the Web all of them depend on statistical mechanisms.

What conclusions can be drawn from my analyses of the statistical style of reasoning in this section? I would argue that statistical reasoning forms the basis of ES^A and of many epistemic practices in the socio-technical epistemic systems of the Web2.0. Taking into account the performative nature of styles, their power of self-authentication, it can reasonably be assumed that the sheer ubiquity and pervasiveness of systems employing statistical mechanisms has led to a further strengthening of an already dominant mode of reasoning. We have to keep in mind, however, that there are other forms of reasoning, other types of epistemic sociality that may be as valid as ES^A – and superior to ES^A depending on the epistemic tasks and goals at hand. ES^A has its merits and accounts, such as Ayres "Super Crunchers" deliver valid arguments for these merits. Clearly, given that epistemology strives for identifying mechanisms that are epistemically beneficial, such analyses are relevant for any epistemology comparing different socio-epistemic mechanisms of decision making and predictions.

Yet there are limits to this type of epistemic sociality. ES^A might be superior for certain epistemic tasks, but it is clearly not suited for others, as has been argued in the previous section. Statistical reasoning can only be applied to quantifiable questions. Some questions may not – or at least not as easily or usefully be quantified. In other words, “[i]f you can’t measure what you’re trying to maximize, you’re not going to be able to rely on data-driven decisions” (Ayres 2007: 150). A crucial question therefore is what gets lost when a statistical style of reasoning becomes too dominant: Which questions can be asked, which can be answered by employing ES^A? Which questions might disappear? If one type of epistemic sociality becomes too dominant, there is a danger that questions which cannot be answered with the mechanisms offered may be rendered irrelevant.

Finally, while statistical procedures may be superior methods for decision making and predictions, this superiority is bound to certain prerequisites concerning independence and diversity outlined above. The message is “garbage-in, garbage-out”: if you do not have good data, no statistical test will tell you anything useful. What it true for the data, is also true for the hypotheses and the assumptions that have informed a study. If they are faulty, the use of statistical procedures is even more problematic, as even Ayres admits: “[s]tatistical analysis casts a patina of scientific integrity over a study that can obscure the misuse of mistaken assumptions” (Ayres 2007: 187). Given the high salience of statistical data and the trust we place in numbers (Porter 1995), we have to make very sure that the data and mechanisms we employ are valid. Otherwise, we may place trust into processes and procedures that are not trustworthy, which brings us to the topic of trust, authority and transparency.

10.4.3 ES^A: Procedural Trust, Algorithmic Authority & The Need for Transparency

In the previous chapter I have outlined the relationship between trust and knowledge and how it figures for ES^I. In this chapter, I argue that while trust is as crucial for aggregational epistemic practices as it was for integrative, the *forms and loci of trust differ* between these two types of epistemic sociality. Moreover, while the basic relationship between trust, reputation, trustworthiness, and authority is the same for ES^A and ES^I, ES^A is characterized by the emergence of a new type of authority: algorithmic authority. I conclude this section with a request for transparency in socio-technical epistemic systems.

To recapitulate: I have argued that trust is indispensable for epistemic practices and thus for knowledge. Without knowledge there can only be blind trust, without trust, there can be no knowledge. I adopted a definition of trust as an attitude towards people, information or mechanisms we hope will be trustworthy. Yet as the word *hope* indicates, we can never be sure that what we trust really will prove to have been trustworthy in the end (Mcleod 2006). Since trust is not being certain, trusting by definition contains the risk of being let down, the risk of having placed trust wrongly. Hence a crucial question concerns the conditions under which trust is warranted. The question of warrant relates to the fact that we do not - and should not - distribute trust evenly over all epistemic agents, content and processes. Rather, we attribute differential amounts of epistemic *authority* to them, depending on how *trustworthy* we consider them to be. Authority implies epistemic trust, by trusting someone to know, we attribute authority to this person with respect to the issue at hand. When we trust others to know, we delegate power to them and make ourselves vulnerable to being let down in case they are dishonest or incompetent.[173]

While trust is an attitude of a trusting epistemic agent to other agents, content or processes, trustworthiness is a characteristic of those who are trusted. That is a crucial, but often neglected difference in debates around trust on the Web and elsewhere. Only because of this difference it is possible to talk about misplaced trust: agents, content of processes that have been considered trustworthy turn out to be untrustworthy, hence trust was ill-placed. Or the other way around: a trustworthy epistemic agent may be denied trust. Both cases are epistemically detrimental. In the first case error may spread, in the second case, growth or spread of knowledge may be hampered.

As was the case for ES^I, we assess epistemic trustworthiness by using various *proxies*. We may assess the trustworthiness of those who provided us information by using their status or their institutional affiliation as proxies. Based on this assessment of trustworthiness we attribute or deny authority to epistemic agents or content. Hence we use reputational cues, some of which are more valid than others. Yet since proxies are heuristics to assess the trustworthiness of others, they are fallible and need to be scrutinized themselves. Two types of errors can occur: we can trust when we should not or we can distrust when we should have trusted: these are the α-error and the β-error of using reputational cues to assess trustworthiness of epistemic agents.

[173] For further information on the issues raised here, please confer also the previous chapter.

In cases when the provider of epistemic content is not known, we may use knowledge about the *process* of its creation to assess whether or not, resp. to what extent we can reasonably trust this content. On the Web this form of trust, trust in processes, is quite frequent, because the providers of content often are not known or because we do not have indicators to assess their trustworthiness. I label this form of trust *procedural trust*. In the previous chapter Wikipedia is already an example of procedural trust. People trust or distrust the information provided on Wikipedia because they trust or distrust the epistemic mechanism underlying Wikipedia. Trusting Wikipedia differs from trust in specific human epistemic agents. Indeed, not only is the person providing content often not known. Without reference to the history page it is also not possible to find out which agent is responsible for which part of the content. Different tools have been developed to make these processes transparent, the WikiScanner as an example to make potential bias due to interests visible, WikiDashboard as an example to make editing patterns visible. If procedural trust is already at work for Wikipedia, this is even more so when it comes to different systems employing aggregational mechanisms.

With reference to Wikipedia, but also to Google and Twitter, Shirky notes that people seem to "trust new classes of aggregators and filters" (Shirky 2009) and calls this tendency *algorithmic authority*. He explains the concept as follows: If someone tells me that Khotyn is a town in Moldova, on which grounds do I trust or distrust this information? In line with my previous analyses on the relevance of social proxies for the prognosis of trustworthiness and the attribution of authority, Shirky notes that we may make use of formal or informal social judgments. An example of a formal social judgment may be a university certificate in geography; an informal one may be that other people have vouched for the trustworthiness, that they trust him. These two examples are reminiscent of Kitcher's distinction of direct and indirect ways of calibrating earned and unearned authority (Kitcher 1993: 314ff).

Since Khotyn is not in Moldova, but in Ukraine, it would have been an example of misplaced trust, if we had trusted this person. Yet this false information was also to be found in the Encyclopedia Britannica until it was corrected. Is there a difference between being wrong by trusting some person you do not know on the Internet and the Encyclopedia Britannica, Shirky asks? There is a difference: because Encyclopedia Britannica is considered to be an *authoritative source* for geographic information, one would feel less silly if it turns out that one has misplaced trust in it.

Shirky concludes that "[a]uthority thus performs a dual function; looking to authorities is a way of increasing the likelihood of being right, and of reducing the penalty for being wrong. An authoritative source isn't just a source you trust; it's a source you and other members of your reference group trust together. [...] It's impossible to be right all the time, but it's much better to be wrong on good authority than otherwise, because if you're wrong on good authority, it's not your fault" (Shirky 2009). This quote stresses a crucial aspect of attributing authority: its socio-epistemic nature. Indeed, Shirky notes: "The social characteristic of deciding who to trust is a key feature of authority — were you to say 'I have it on good authority that Khotyn is a town in Moldova', you'd be saying that you trust me to know and disclose that information accurately, not just because you trust me, but because some other group has vouched, formally or informally, for my trustworthiness" (Shirky 2009). Epistemic authority is shared epistemic trust.

So far this has been a general characterization of authority as a socio-epistemic construct, which corresponds to the analyses on authority delivered in the previous chapter. The crucial innovation, the change that Shirky observes, however, refers to the emergence on a new form of authority: algorithmic authority. This form of authority differs from authority ascribed to epistemic agents. He describes the formation of algorithmic authority as follows (Shirky 2009): First material from multiple sources, which themselves are not vetted for their trustworthiness, has to be combined in a process that is not supervised by some editor before being made available. Second, this mode of providing information must lead to good results so that people start to trust it. Yet only once people realize that others trust this source as well, it turns into an authoritative source of information for this community of people. This means that the attribution of authority is a decidedly *socio*-epistemic process.

Referring back to my previous conclusions on the relationship between trust, trustworthiness and authority in ES^A, this development implies that on the Web, in applications that employ ES^A, people start trusting aggregational mechanisms because they a) delivered some good results (direct verification) and b) others trust these mechanisms as well (indirect calibration of authority). They place trust in certain processes and mechanisms, a form of trust which I have labeled procedural trust to distinguish it from trust in epistemic agents. In trusting these processes, they collectively attribute authority to aggregational mechanisms. Trust is placed in the aggregational mechanism of closure and not in the individual epistemic agents. Indeed, Shirky seems to directly refer to my considerations concerning the statistical rationale behind ES^A when he concludes:

> "But the core of the idea is this: algorithmic authority handles the "Garbage In, Garbage Out" problem by accepting the garbage as an input, rather than trying to clean the data first; it provides the output to the end user without any human supervisor checking it at the penultimate step; and these processes are eroding the previous institutional monopoly on the kind of authority we are used to in a number of public spheres [...]" (Shirky 2009).

What conclusions can be drawn from Shirky's "speculative post on the idea of algorithmic authority" [...]" (Shirky 2009)? First of all, I would argue that it is a key characteristic of ES^A that there is a concentration of trust on the aggregation mechanism. While for ES^I about equal trust was needed in all components of the epistemic process, trust in ES^A is much more skewed. Below is the model which I have sketched to indicate all the instances and processes which need to be trusted within ES^I. Trust needs to be placed in epistemic agents, in their interaction, in the integration processes as well as in knowledge as a result of these epistemic processes if it is to be fed back into a process of knowledge creation.

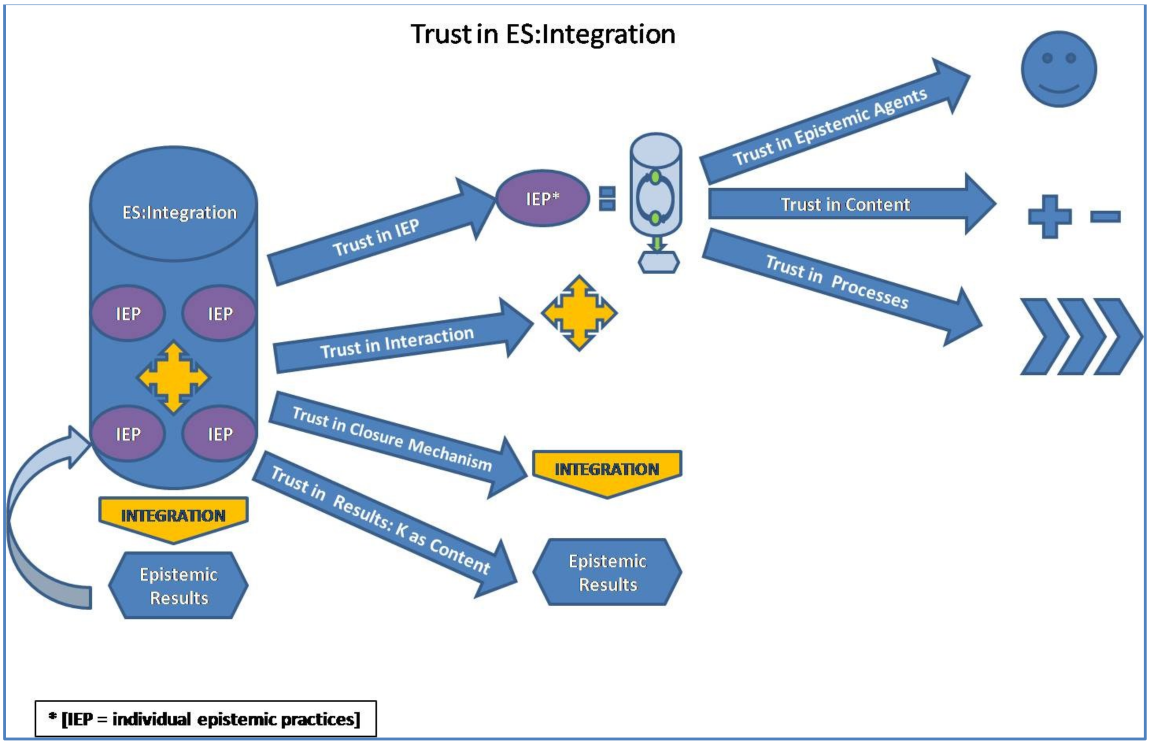

Figure 17: Trust in ES:Integration

This is different for aggregational mechanisms as I have indicated in the next picture. Here much less trust is needed in the individual agents. One needs to place some trust in the collectivity of epistemic agents, namely that they are diverse and independent, but the competence and honesty of the individual epistemic agents are of lesser concern. The reason for this is that according to the ES^A rationale, error will cancel itself out as long as there is a large enough number of epistemic agents who are independent and diverse. Hence, less trust needs to be placed in the individual epistemic agents, while more trust needs to be placed in the aggregational mechanisms. This trust in the methods of aggregation is what Clay Shirky refers to as algorithmic authority.

However, the problem with this algorithmic authority is that the underlying heuristics and techniques frequently are inaccessible or incomprehensible for users. Hence, they are in the very vulnerable position of having to rely on sources whose *heuristics* they cannot control and which *biases* they cannot assess. Since many algorithms, such as Flickr's *interestingness,* Google's PageRank, or the algorithms underlying different recommender systems, are either unknown or imcomprehensible to most if not all users, users of such systems cannot be reponsible knowers, because they do not even understand the processes upon which they rely. This issue is of particular concern in reputational algorithms, because of the effects these algorithms has not only on epistemic content, but also on epistemic agents. The general pros and cons of using reputation for epistemic purpose have been outlined in the previous chapter. The specificity of reputation in ES^A lies in the fact that reputation is formalized through algorithms and as such embedded – and often rendered invisible – within socio-technical epistemic systems. If reputational tools on the Web have to be scrutinized by epistemically responsible users who do not want to accept too naïvely the outcome of a process they do not control, the basics of these tools need to be made transparent. That is, the responsibility to scrutinize socio-technical epistemic systems can only be fulfilled if it is possible to access the underlying techniques and heuristics, the algorithms in the first place. Moreover, it must be possible to make sense of them. This implies a duty for designers of such systems to make the algorithms that they use transparent in order to empower users to become responsible knowers. One example of how such transparency could be implemented is sketched in the next section.

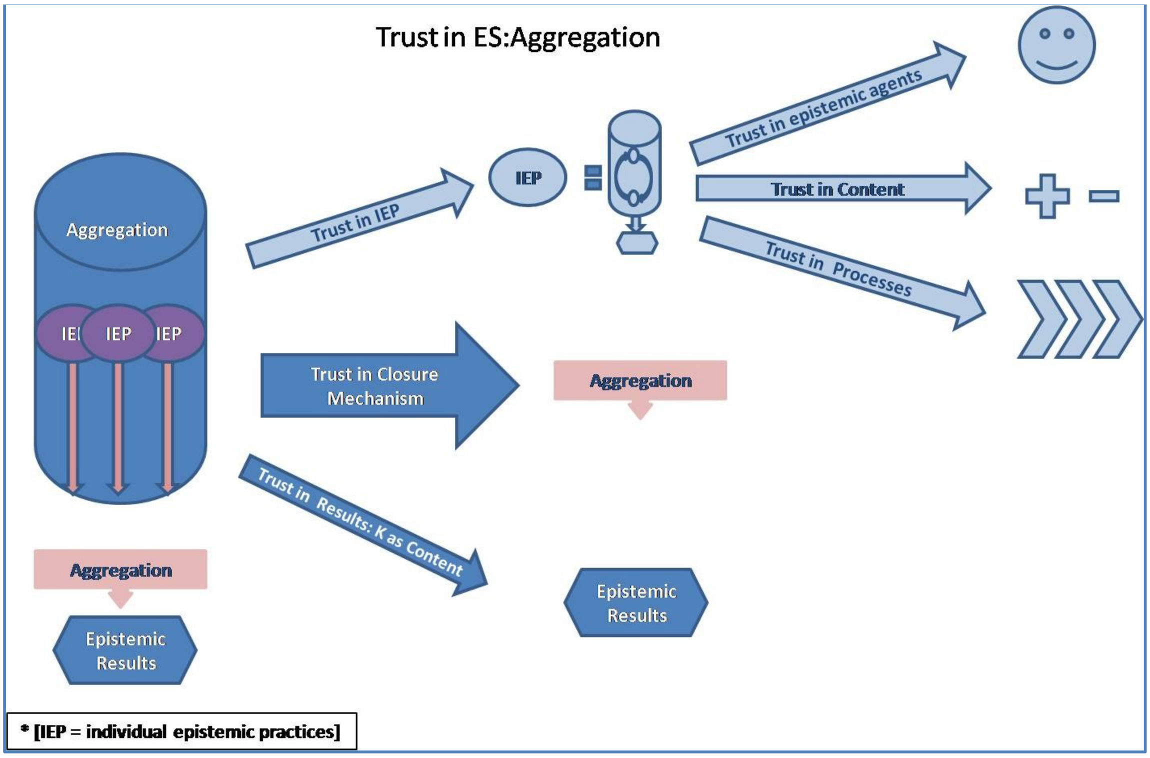

Figure 18: Trust in ES:Aggregation

10.5 Implications for the Analysis and Design of Socio-Technical Epistemic Systems: An example

10.5.1 Trust, Authority and Transparency in Trust-Aware Recommender Systems

Trust-aware recommender systems as a specific type of recommender systems (RSs) have already been briefly introduced in Chapter 2. In this chapter I want to return to these systems to elucidate the relationship between knowledge, trust, authority and transparency in ES^A. A brief reminder: besides trust-based systems, there are different types of classical RSs: content-based, collaborative-filtering as well as different hybrid systems. They have in common that they all serve the purpose of suggesting content of interest to users. From a *socio*-epistemological perspective, collaborative-filtering systems are of greater interest than content-based systems, because here the recommendations are based on judgments provided by the community of users, by multiple epistemic agents of a socio-technical epistemic system. Hence, collaborative-filtering mechanisms are an example of ES^A.

Classical RS techniques have several shortcomings, including the so-called cold start problem, i.e. the difficulty to generate recommendations for new users (Massa and Bhattacharjee 2004). When a new user enters a system, the system does not “know” anything about this new user and this ignorance makes it difficult to generate appropriate recommendations for her. To counteract this problem, traditionally, new users have been asked to rate a few items so that the system can “learn” something about the user in order to provide personalized information on interesting items for her. However, especially in large databases necessary correlations are scarce and thus, this procedure often turns out to be quite ineffective. In consequence, algorithms have been developed that bootstrap the system not by feeding in *judgements on content*, but *judgements on other users*. Systems employing such algorithms are trust-based recommender system. Numerous trust-based RS have been proposed, such as Advogato by Levien (Levien and Aiken 1998), Appleseed by Ziegler and Lausen (Ziegler and Lausen 2004), MoleTrust by Massa and Avesani (Avesani, Massa et al. 2005, Massa and Bhattacharjee 2004) and TidalTrust by Jennifer Golbeck (Golbeck 2006; Golbeck and Hendler 2006). [174]

In the following I portray the works of Paolo Avesani, Bobby Bhattachcharjee and Paolo Massa as an exemplar. I have chosen their model, because the authors highlight several issues of high socio-epistemological interest. These issues include the epistemic relevance of trust, the relationship between trust and authority, and finally the relationship between trust, different forms of bias, and (in-)transparency. Massa & Bhattachasjee have developed an algorithm for “Trust-aware Recommender Systems”, arguing that the cold-start problem can be solved by implementing a notion of trust between users into the system (Massa and Bhattacharjee 2004). The difference between traditional RSs and trust-aware RSs is quite simple: “While traditional RSs exploit only ratings provided by users about items, Trust-aware Recommender Systems let the user express also trust statements, i.e. their subjective opinions about the usefulness of other users” (Massa and Avesani 2006). This seemingly minor change proves to be highly effective to remedy the cold start problem because “it is able to exploit trust propagation over the trust network by means of a trust metric” (Massa and Avesani 2006).

[174] For a review on models of computational trust and reputation please confer (Sabater and Sierra 2005) as well as (Wang and Vassileva 2007).

With respect to the relationship between knowledge, trust and reputation this example implies that assessing one's peer and placing differential amounts of trust in them is epistemically useful, because it enables better predictions within RS. This means that reputational mechanisms, mechanisms in which trust is placed in some agents but not in others, have epistemic merits. However, a second point is interesting and it concerns our trust in the algorithms themselves. As Shirky has noted, there seems to be an increasing willingness to attribute epistemic authority to algorithms (Shirky 2009). While we have seen the epistemic merits of aggregational mechanisms throughout this chapter, the limits of these mechanisms as well as the problems that occur, if certain prerequisites are not met, have also been outlined. Hence, to be epistemically responsible users of such mechanisms or systems that employ such mechanisms, we would need to make sure that the prerequisites are met and that appropriate aggregational mechanisms are employed. Only then can we place trust responsibly. Unfortunately, we often trust these algorithms blindly and we frequently are even forced to, because algorithms are hidden within a system. In most cases we are not aware how they work and we cannot assess their impact on the information we receive. In other words: algorithms are *black-boxed*. Once a system is released the way it works seems without alternatives - and the negotiations and decisions that have taken place in its development become invisible (Pinch and Bijker 1987).

To get an idea of the relevance of such invisible decisions it is instructive to compare the effects different algorithms have when being applied to the same data set. As an example, I briefly compare two types of trust metrics: local and global trust metrics (Massa and Avesani 2007). Trust metrics are techniques for answering questions such as "Should I trust this person?" in virtual communities. Hence they are means to estimate the trustworthiness of information providers. Depending on the metric, on the algorithm you choose, you will receive different values of trustworthiness for each user. That means that even if a distinct value of a user existed (her "true" trustworthiness), the predictions of different metrics would be dispersed around this true value. This is a basic effect of using statistical methods for predictions. However, the problem may be even more profound. What if there is no such thing as a true value of trustworthiness of a user? What if the user A's trustworthiness was different for user B and user C? This is the rationale behind local trust metrics: while "[g]lobal trust metrics assign to a given user a unique trust score, the same independently of the user that is evaluating the other user’s trustworthiness […], a local trust metric provides a personalized trust score that depends on the point of view of the evaluating user” (Massa and Avesani 2007: 40). From the perspective of local trust metrics, the trustworthiness of a user is not an intrinsic value of that user, but lies in the eyes of the beholder. This difference should be reminiscent of certain debates in epistemology: while some argue that knowledge is always situated and contingent, others adhere to universal knowledge claims.

Moreover, different algorithms have differential effects on users as becomes obvious for *controversial users*. Controversial users are users who are trusted and distrusted by many other users of the system. By using a global trust metric that calculates an average trust value for each user, these users would simply be “averaged out”. They would be rendered irrelevant by receiving the weight “zero”. It is of epistemological and ethical interest that controversial users are valued very differently in these two different metrics. Local trust metrics explicitly stress and appreciate the individuality and situatedness of every trust statement and state that controversial users by definition do not have a global trust value for the whole community. By contrast, global trust metrics suggest a fictitious consensus between users by calculating an averaged trust value for each user. Through this process, the controversial user is rendered “unreliable” and gets statistically eliminated.

These different types of trust metrics do not only have different underlying assumptions about the value of those users and about deviation from the mean - or norm - more generally. They also have an impact on the information one receives and whose opinions are included. They might even retroact on cultural and societal values on how to deal with minority views. Averaging out controversial users by means of statistics has a similar effect as other mechanisms of “sorting out” (Bowker and Star 1999) and silencing: they exclude those from participation that deviate too much from the norms or do not fit in ready-made categories. To conclude, local trust and global trust metric are based on very different epistemological and ethical premises. Not only do they differ with respect to their stance towards universality versus situatedness of trustworthiness. They also value different user types differently and have differential effects on the importance of such users. In Nissenbaum and Friedman’s terms, RSs employing global trust metrics entail bias: they *systematically* and *unfairly discriminate* against certain individuals, namely controversial users, in favor of others, namely all non-controversial users (Friedman and Nissenbaum 1997: 23).[175] In Friedman and Nissenbaum's systematization, this would be *technical* bias, because it arises within the process of technology design, when the developers decide upon the algorithms they choose to employ. However, I do not think it is entirely clear that this technical bias can be completely separated from *pre-existing* and *emergent* biases. Given the fact that global and local trust metrics correspond nicely to different epistemological positions, it cannot be precluded that at the very least implicit societal values are being inscribed into RSs by opting for one algorithm and not the other. Moreover, emergent bias can surely evolve and hence can also not be excluded. This is not to imply that local trust metrics are generally preferable or that they cannot entail biases. Depending on the purpose global trust-metrics may indeed be useful as shall be outlined below.

What conclusions can be drawn from my analyses? We have seen that different algorithms in trust-aware RSs lead to different recommendations and affect users of the system differently. Moreover, these processes are usually black-boxed and therefore not assessed critically. What if we made such algorithms visible? What if we left it up to the user to decide whether he prefers to use local or global trust metrics? All that would be needed to increase the transparency of RSs and to empower its users is a dual search-button and a way to visualize the differences between the different algorithms. In the following I sketch a simple, yet epistemologically and ethically relevant extension to trust-aware RSs. I have labeled my thought experiment "MyChoice". I chose this name for two reasons. First it is supposed to indicate that epistemologically and ethically relevant choices are constantly being made in the course of developing software. Thus, by the label My*Choice*, users should be made aware that these decisions are built into software and have certain effects. Secondly, *My*Choice is a tool that empowers users to make informed choices of their own where it is normally the programmer who has decided for them.

Basically, MyChoice has two distinct features. One is a dual search button, by which you can choose between two different trust metrics to generate recommendations. The labels that I have chosen are "Search... personalized for me" versus "Search... the golden mean". They correspond to the local and global trust metrics as described in (Massa and Avesani 2007). One can set one of the trust metrics as one's default. However, by

[175] Please confer Chapter 3 for further information.

clicking on one of the two buttons one can change this for each new query. The second feature is an "open-eye"-button. If you click on this button, the differences between the two trust metrics will be visualized.[176]

As noted before, there are two differences: different recommendations and different effects on the users of the system, who are given different weights. The different recommendations could be indicated simply by listing them next to each other. Users would see the recommendations based on global trust metrics on the right half of the page and those based on local trust metrics on the left. The differences would be visible at a glance. And by seeing *two different lists of recommendations* users would become aware of the effects algorithms have on the information they obtain – effects that they were formerly not aware of, simply because they were not visible.

MyChoice would also empower users to decide which algorithm works better for them. The dual search button enables users to make more informed decisions about the information she wants to receive. For instance, it would be possible to switch between the global and the local metric depending on the context. Users might opt for the more situated, local option when looking for movie recommendations, but for the rather universalist, global one when they want to learn something basic about computing, statistics or gardening. In fields where users are novices or for certain reasons more interested in mainstream recommendations, they might press the "Golden Mean"-button. For other questions in fields where they are either more knowledgeable or interested in certain niches, they might prefer the "Personalized For Me"-search. In the end this decision is up to the user. Yet simply by using MyChoice, users will learn about the functioning and the consequences of different metrics and algorithms and this will have positive epistemological and ethical consequences by making them more responsible and reflective knowers.

As was shown in the case of the controversial user, different algorithms also value users differently. To make these differential effects of local and global trust metrics on users visible, one might use social network graphics. One could again place two social network graphics next to each other and indicate the different weights of users by different color codes. Depending on the algorithm, the color of users may change,

[176] I have drafted an interface for MyChoice below. I have superimposed my draft on a screenshot of the website Epinions.com, because the analyses by Massa and Avesani (2007) were based on a data set from this Website: http://www.epinions.com/ [date of access: 30.11.2009].

indicating a differential treatment by the two algorithms. Highlighting those groups of users who are most affected by changes in trust metrics will render the discriminative consequences of different metrics visible. Moreover, this information about different users and user types might not only be illuminative for the users, but also for developers trying to improve their metrics and algorithms.

What would be the utility of a tool such as MyChoice? First of all, such a system empowers the user, because it offers her the chance to decide upon which algorithms she prefers for different purposes. Moreover, people using this system would become much more aware of how implicit assumptions and values of the programmers are inscribed into technology. Users would realize the impact of certain programming decisions on the retrieval of information and on different user groups. This effect would already be triggered by the dual search option, but it would be amended by the visualization of the different search mechanism and differences of retrieved information.

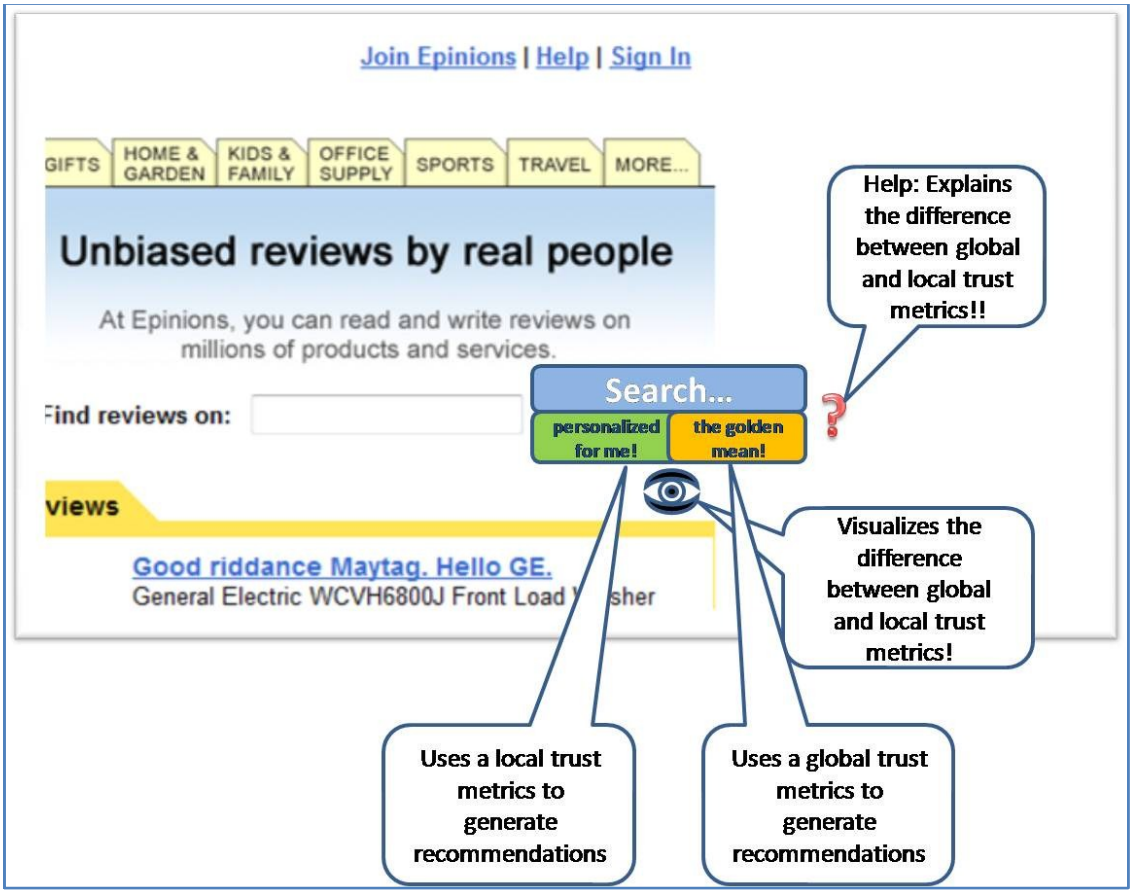

Figure 19: Draft of an Interface for MyChoice[177]

[177] Superimposed on a screenshot of the Epinions.com website.

10.6 Conclusions on ES:Aggregation

In this chapter I have outlined another type of epistemic sociality by which multiple epistemic agents together achieve epistemic goals: ES:Aggregation. I have argued that this mechanism differs from ES^I and ES^S and describes a distinct mode of epistemic sociality. It differs from ES:Integration because the agents do not need to communicate, coordinate their activities or even interact. Indeed such interaction can be harmful, because their judgments need to be *independent* before being aggregated. And it differs from ES:Selection, because no epistemic agent alone can achieve the desired results - without the aggregation process, there is no epistemic result. The input from various epistemic agents needs to be aggregated for this type of epistemic sociality to come to the fore.

I have discussed the specificities of ES^A as compared to ES^I and ES^S with respect to several key topics. Yet one major distinction which was emphasized throughout this chapter referred to the fact that aggregational mechanisms can be applied to judgments about *epistemic content* or to *epistemic agents.* If both happens – as often is the case on the Web - one could speak of “Epistemic Sociality^2: Aggregation^2”, because the mechanisms are both social and aggregational in a dual sense.

This means that the basic mechanism of ES^A consists in the *statistical aggregation of judgments from epistemic agents that are independent and diverse*. It is social because information provided by multiple epistemic agents is used in aggregated form to provide a collective result. If this mechanism is applied to epistemic agents, this process becomes doubly social: not only is information aggregated – it is *social information* (Lehrer and Wagner 1981) that is being aggregated. This social information, information about one’s peers is aggregated to assess the *reputation of epistemic agents*, which then can serve as a weight for their judgments in the further process of judgment aggregation. This is where it becomes doubly aggregational: first aggregation is used to achieve consensual weights of trustworthiness of epistemic agents, then aggregation is used to decide upon topic matters weighted and the judgments of epistemic agents are weighted by their reputation.

Therefore, conclusions have to be drawn about two issues:

1) The basic mechanism of aggregating judgments from multiple epistemic agents:

This aspect refers to the pros and cons of *statistical reasoning*. We have to ask to what extent such aggregational mechanisms are epistemically useful and where potential dangers lie. Here two issues are crucial. One concerns the problems that occur when either the prerequisites for the data are not met or the algorithms are not useful. The second concerns the issue of algorithmic authority, i.e. the question of how much trust we should put into algorithms and how much transparency is needed to be justified in trusting algorithms.

2) The application of aggregational mechanisms to social information:
This question refers to the use and misuse of reputational cues to weight the judgments of epistemic agents. First of all, the weights that are ascribed to others, even if numbers look so convincing, are initially based on idiosyncratic assessments and subjective judgments of peers. This human subjectivity is decisive at various stages: the initial input values for reputational mechanisms in Lehrer's model, the subjective judgments in recommender systems, the subjective decisions of programmers for and against certain formalizations and algorithms. This implies that pre-existent as well as technical and emergent bias become relevant. A crucial danger resides in the possibility of epistemic injustice: the problems that occur when inappropriate proxies are used to assess the trustworthiness of epistemic agents. The problem of epistemic injustice is not unique to ES^A, it exists for all types of epistemic sociality. However, the problem with using reputational measures in systems employing ES^A consists in the fact that these proxies are rendered invisible by being inscribed into technology, by being hardwired into systems and by being blackboxed if the algorithms and the input data are not made transparent. Hence, a crucial requirement for reputational mechanisms in socio-technical epistemic systems concerns their transparency.

As was the case for the previous chapter, I conclude my considerations on this type of epistemic sociality with a summary of the main criteria that should be fulfilled to ensure the proper functioning of systems employing ES^A. Again, these criteria for systems analyses and systems design are rather generic and need be further specified.

Setting the Parameters for ES:Aggregation	

Epistemic Tasks	• The tasks that individual epistemic agents conduct tend to be smaller and more embedded in other practices than is the case for ES^I and ES^S. Hence motivation is less of a concern.
	• With respect to the overall epistemic tasks of content production versus content evaluation, ES^A is mostly rather employed for content evaluation, filtering and selection.
	• However, ES^A is useful and used for the creation of two specific types of epistemic content:
	  ○ second-order knowledge such as bottom-up classification systems (folksonomies).
	  ○ collective predictions and collective decisions.
	• One possibility to ensure independence on the level of the task structure is to offer tasks simultaneously to multiple agents.
	• ES^A is well suited for epistemic tasks with singular, quantitative outcome (single values, orders or likelihoods).

Numbers, Diversity, Independence & a Statistical Style of Reasoning	• Human epistemic agents should be numerous to enable a statistical averaging out of error. • They should be diverse to avoid bias. Diversity here follows a statistical rationale of ensuring statistical variance. • They should act independently to avoid information cascades and the spreading of error. • They need to have some private information, some knowledge about the issue at hand. The rationale of cancelling out error to be left with information depends on the existence of some information in the first place. • Agents need to be less motivated than for ES^I, because individual tasks tend to be either smaller (e.g. a vote on some content) or embedded into practices done for oneself (e.g. tagging something on Delicious). For prediction markets there can be additional financial incentives.

	• Be aware of the limits of statistical reasoning! • Care about the data input and remember: Garbage in-Garbage out! • Beware of patina of scientific integrity through statistics that hide faulty assumptions, data or methods!
Procedural Trust, Algorithmic Authority & Transparency	• Be aware of and try to avoid different types of bias (e.g. statistical, pre-existing, technical and emergent bias). • Raise awareness about unconscious and implicit use of reputational cues to assess epistemic trustworthiness of epistemic agents and about the danger of epistemic injustices, i.e. the use of invalid social proxies to assess epistemic trustworthiness. • Be transparent! Make functionalities, implicit assumptions, design decisions & effects of different algorithms visible.

Table 6: Setting the Parameters for ES:Aggregation

11 Epistemic Sociality 3: Closing Socio-Epistemic Processes Through Selection

11.1 Epistemic Sociality: Selection (ES^S)

In this chapter I outline my third and final type of epistemic sociality: ES: Selection. It is a distinct form of epistemic sociality, with its own rationale, benefits and problems. Instead of *integrating* or *aggregation* epistemic results provided by different epistemic agents via the various mechanisms portrayed in the last two chapters, ES^S *selects* among those epistemic results. Thus, as a closure mechanism, ES^S differs from ES^A and ES^I, because the individual results among which one chooses *do not have to be re-combined* in one way or another. However, ES^S is particularly suited to elucidate one aspect of my three types of epistemic sociality which I have so far brushed under the carpet: *the possibility to combine and nest different types of epistemic sociality in socio-technical epistemic system and processes*. How and when such nesting can take place in ES^S becomes obvious when taking a closer look at the temporal structure of ES^S-type socio-epistemic processes. For selection to be a *socio-epistemic closure mechanism*, at first epistemic work needs to be distributed over multiple epistemic agents. In contrast to ES^I and ES^A, these epistemic agents however, aim at fulfilling the *same* epistemic goal. They may employ different methods or follow different approaches, but the goal is the same. This is the first moment in time when types of epistemic sociality can be nested, because to reach an epistemic goal, different forms of epistemic sociality can be employed: Agents can collaborate with each other to fulfill the epistemic task (ES^I), they can use aggregational mechanisms (ES^A), or they can individually pursue different tasks. All these options are possible as indicated by the picture below.

Once these singular epistemic results, obtained through different individual or socio-epistemic mechanisms, are presented, a judgment has to be formed on how to distinguish them. Here again, different integrative or aggregational mechanisms can be used, and therefore this is the second moment in which different types of epistemic sociality can be combined. In principle there are three generic forms of judgment formation.

a) dictatorial: one epistemic agent alone judges epistemic content or epistemic agents provided by others.

b) aggregational (ES^A): multiple epistemic agents *vote* on content or other agents, these votes are aggregated, different possibilities of weighted and unweighted aggregation are possible.

c) integrative (ES^I): multiple epistemic agents *deliberate* on content or other agents, the individual judgments are integrated, different possibilities of weighted and unweighted integration are possible.

Finally, the selection has to be made. It could be *random*, i.e. without prior differential assessment of the submission – or it can be based on any of the modes of judgment formation proposed above. It can also take different forms: *picking one*; *picking some,* i.e. drawing a line and distinguishing between accepted and unaccepted epistemic content (e.g. selecting submissions for a conference) or agents (e.g. selection fellowship applications); *ordering or ranking* content or agents.

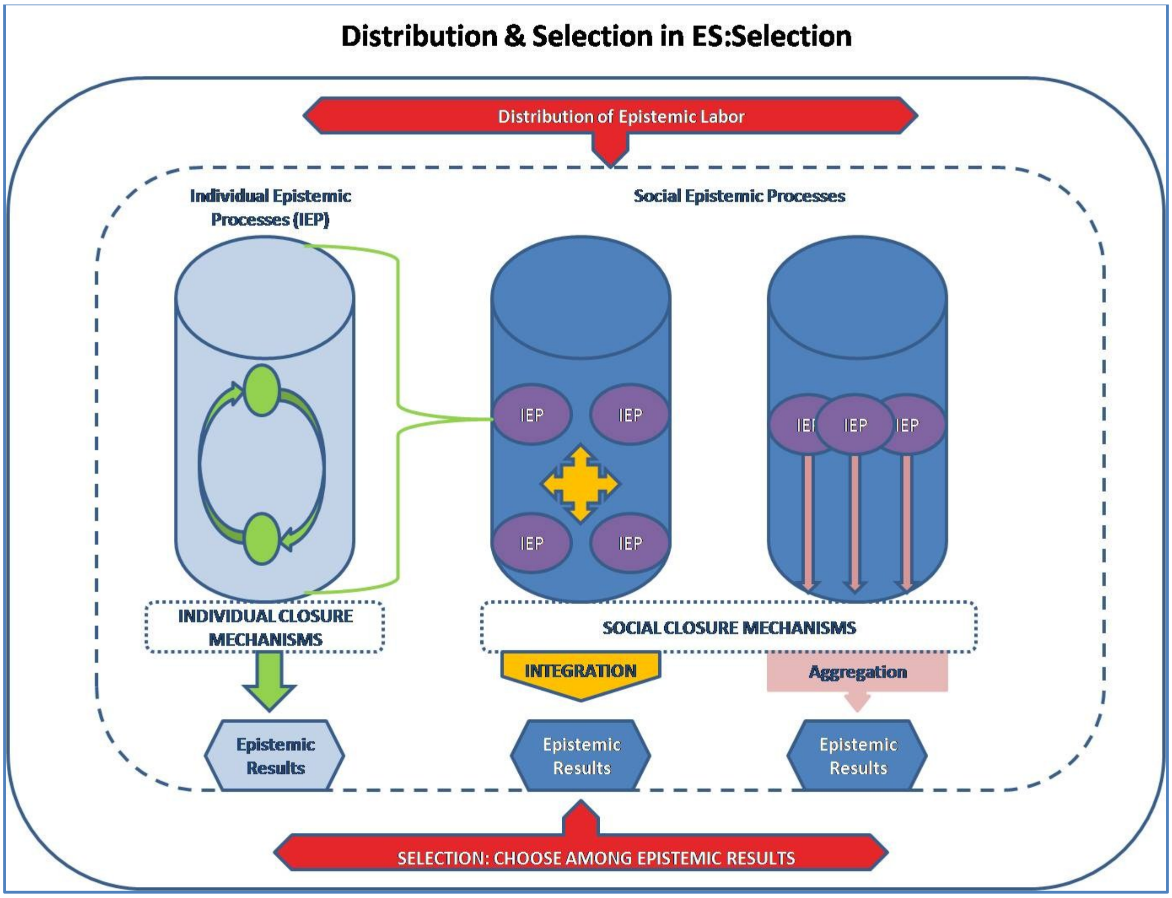

Figure 20: Distribution and Selection for ES:Selection

To understand ES^S and the possible nesting of types of epistemic sociality, imagine the following situation. A price is announced for the best introductory book on social epistemology. Numerous social epistemologists submit their books. Now the choice has to be made among these submissions and different options are available to take this decision. It could be that the donor of the price money wants to choose the winner herself. Whether or not this is an advisable and epistemically beneficial strategy may depend on certain characteristics of the donor: Is she competent? Biased? Epistemically trustworthy? It is also possible that a committee of different people discusses and consents upon the selection. Here not only the characteristics of the individuals involved in the process are of interest, but also how they come up with their collective decision. Which mechanisms are chosen to integrate the different judgments and opinions into the collective decision? Is everyone's point of view taken into account? Does everyone's judgment have the same impact on the outcome or are different agents dominant? Is this dominance justified because the agent is considered to be the most competent? Or is he simply the loudest? Finally, it is also possible to have a random mass of people vote on the different proposals and pick the one with the most votes. Here questions arise concerning the aggregation mechanism: Is a weighted or unweighted procedure employed, if so for what reasons, etc.

In processes involving multiple epistemic agents who either deliberate or vote, integrative or aggregational mechanisms are employed to obtain a collective judgment upon which the selection can be based. This is what I mean by saying that ES^S can be based upon ES^I or ES^A. In these two examples ES^S is socio-epistemic, because it employs social mechanisms of combining epistemic judgments of multiple agents in ways outlined in the previous two chapters. Due to this possible nesting of types of epistemic sociality, all topics, merits and problems raised in the previous two chapters become relevant for ES^S whenever integrative or aggregational mechanisms are employed. In order not to repeat myself, I focus only on some *additional* issues that are specific to ES^S in this chapter.

Concerning the sociality of ES^S, it is crucial to understand that even the first case, the dictatorial selection has a socio-epistemic component. It is social in the sense of offering a *macro-perspective* on the distribution of epistemic labor within a community. If it is possible to choose between different epistemic results for the *same question*, different agents must have worked – possibly in different ways - to come up with these results. Selection is the closure mechanisms, but the process that proceeds this closure is *distributed epistemic labor*. This distribution can take the forms characterized in the previous two chapters for ES^I and ES^A, but work can also be distributed over different *individuals* who are in competition with each other to fulfill the same task. The crucial point is that numerous epistemic agents, either individually or organized in forms characteristic of ES^I or ES^A pursue *the same epistemic goal by different means*.[178] This process of letting different agents try our different methods and approaches, hoping that one of them finds the solution is epistemically advisable, because before the problem is solved, one does not know which strategy will succeed. Therefore if different strategies are pursued in parallel, the solution to a problem is expected to be found earlier. This is the perspective of Kitcher's philosopher-monarch (Kitcher 1993: 305), the perspective of Solomon's *more social epistemology* (Solomon 1994), the macroscopic perspective of science policy: how and by what means to structure an epistemic field so that at least one epistemic agent succeeds in fulfilling an epistemic goal.

11.2 ES^S & Social Epistemology: Diversity and the Distribution of Epistemic Labor in Science

In social epistemology, issues related to ES:Selection have been addressed primarily under two headings: the division of cognitive labor (Solomon 2001, Kitcher 1993) and epistemic diversity (Solomon 2006). Since many social epistemologists are also philosophers of science, their analyses have primarily focused on science. We will see that there are some crucial differences between exploiting ES^S in science and the Web. While it is often stressed that science creates new knowledge, whereas on the Web knowledge is primarily distributed, I argue that this is *not* the crucial difference here. After all, one of the examples of epistemic social software introduced in Chapter 2, Innocentive.com, explicitly encourages the creation of new knowledge. I return to Innocentive.com below. The more relevant difference concerns size: the *number* of available epistemic agents and their *diversity*. Many social epistemologists and philosophers of science have advocated diversity when discussing the optimal

[178] It is possible that for certain epistemic problems, individuals alone can propose answers; singular epistemic agents can provide solutions. Although according to the socio-epistemological position adopted in this thesis, individual epistemic agents alone can never produce knowledge, because knowledge is a social status that depends upon its ascription through a community, individuals can of course submit epistemic content into this process of communal assessment. That even these individual submissions are *social* in the sense of building upon knowledge created by others, etc., has been outlined before and is of course still considered valid. Hence, epistemic content of different forms can be produced by singular agents who are nonetheless related to other epistemic agents in various ways. And for epistemic content to be considered knowledge, is has to be vouched for by a community of knowers bound together by a set of entitlements and commitments (Kusch 2002).

distributions of cognitive labor in science. In this chapter, this question concerns the distribution of cognitive labor over different theories or approaches within a scientific field as opposed to the distribution of tasks within research teams. The latter topic has been addressed in Chapter 9 on ES:Integration. Thus, the focus is more macroscopic and looks at overall fields of research rather than at specific research teams. While the benefits of diversity are widely acknowledged in social epistemology, there is little consensus on whether and if so by which means diversity should be reinforced in science.

In a paper on "Norms of Epistemic Diversity" Miriam Solomon argues that diversity is valuable because it produces dissent which in turn produces four valuable outcomes: 1) worthwhile criticism, 2) division of labor, 3) social distribution of knowledge and 4) creativity (Solomon 2006). To my mind the logic is a bit different. The *social distribution of knowledge* as well as potential increases in creativity are two *reasons to support diversity*, while worthwhile criticism may be a *result*. The division of cognitive labor in turn is a *cause as well as a result* of diversity and dissent. With respect to knowledge, diversity and dissent finally are mutually related: *cognitive diversity*, i.e. different ways of reasoning based on the employment of different perspectives, heuristics, interpretations and predictive models (Page 2007) may lead to dissent; while dissent may lead to *epistemic diversity*, i.e. the pursuit of different epistemic approaches. Here, I use cognitive diversity in Page's sense as referring to different perspectives, heuristics, interpretations and predictive models (Page 2007). Epistemic diversity as understood here encompasses cognitive diversity but goes beyond it by including epistemic practices, such as the knowledge-productive practices described by Longino (Longino 2002c).

In the same article, Solomon also distinguishes three different philosophical stances towards diversity in science (Solomon 2006). The first one is a laissez-faire view, according to which there is enough diversity in science and that for various reasons there should be no intervention into science. Solomon attributes this view amongst others to Philip Kitcher, whose perspective is depicted below. Solomon contrasts this *laissez-faire* or *invisible hand model* with feminist positions that demand explicit encouragement and support of more diversity, which propose specific kinds of reconfiguring the scientific field by supporting non-mainstream research. Solomon refers to Longino (Longino 1990) and Evelyn Fox Keller (Keller 1985) as two proponents of such interventionist models, who argue that more research using complex linear models as opposed to research following hierarchical linear models should be encouraged. While Solomon agrees that there should be more diversity in science, she refrains from making claims about which *type* of diversity might be needed. More specifically, she asserts that "diversity is a blunt epistemic tool" (Solomon 2006: 26). That is, instead of promoting more research with feminist values one should rather *manipulate the distribution of bias* that works in favor or against different theories to make dissent appropriate.

11.2.1 Solomon's Position on Diversity: Manipulate the Vectors

The idea of manipulating bias is rooted in Solomon's own social epistemology, which she has labeled *Social Empiricism*. It was described in detail in Chapter 5. To understand how she relates dissent, diversity, bias, let me recapitulate the main features of her approach (Solomon 2001, Solomon 1994). Solomon's *Social Empiricism* is intended to be a *more social* epistemology (Solomon 1994). Instead of focusing on how research is conducted in teams, Solomon adopts a macroscopic perspective that focuses on the distribution of cognitive labor within research communities. The guiding question is how many, which and for how long different research strategies should be pursued. She argues that consensus is a special case of dissent, namely zero degree of dissent and that it is normatively appropriate only in very few cases. Analyzing case studies from the history of science, she rejects the idea that consensus is intrinsically valuable or the ultimate goal of science. Indeed, premature consensus can hamper science by precluding the pursuance of alternative approaches and according to her case studies, such premature consensus has frequently occurred in history of science. Moreover, there are not only cases in which consensus was premature, but also cases in which it was kept up for too long despite evidence against a consensual theory. Solomon analyses the role various social factors play for consensus formation, retention, and dissolution in science and the effects that these patterns had on the development of science in different fields. Instead of considering social factors as biases, she introduces the term *decision vector* – as a neutral placeholder for various social effects. She distinguishes between empirical and non-empirical decision vectors[179] and argues that not the single vectors, but their

[179] While empirical decision vectors are "[…] causes of preference for theories with empirical success" (Solomon 2001: 56), non-empirical decision vectors lack this connection to empirical success. Examples that Solomon gives for non-empirical decision vectors include social and political (e.g., ideology, peer

distribution affected the welfare or stagnation in different scientific disciplines at various stages in history.

The core of her *Social Empiricism* lies in providing recommendations on when consensus, dissent and the dissolution of consensus should take place. While *consensus* is only appropriate if one theory can account for all empirical evidence, her conditions for *dissent* are necessary to understand her recommendations concerning diversity to be outlined below. According to Solomon's *Social Empiricism*, dissent is appropriate when the following three conditions are met.

> "1. Theories on which there is dissent should each have associated empirical success.
>
> 2. Empirical decision vectors should be equitably distributed, i.e. in proportion to empirical success.
>
> 3. Non-empirical decision vectors should be equally distributed (the same number for each theory" (Solomon 2006: 117f).

She assumes that the first two prerequisites are easily satisfied and that therefore any attempt to increase dissent - and thereby epistemic diversity – must lie in *manipulating the non-empirical decision vectors*. If these are equally distributed, dissent is legitimate and the distribution of research effort over several alternatives in justified. Hence, if we want to raise epistemic diversity, i.e. the pursuance of different approaches, we must make sure that the non-empirical decision vectors, all those biases that are *not* related to the empirical data affect alternative theories equally.

To elucidate her claim, Solomon portrays an example of an unequal distribution of decision vectors in science, which to her mind has hampered diversity. She depicts the "story of a novel scientific theory and its rejection (so far) by both the science and the technology communities" (Solomon 2006: 28). This theory of concern is Luca Turin's theory of smell, according to which smell detectors work by distinguishing the vibrational frequencies of molecules (Solomon 2006: 28). This theory is in opposition to the mainstream theories according to which smell is detected by receptors recognizing the *shape* of odorant molecules. Based on the analysis of a biography of Turin, a book written by Turin about his theory, as well as some of his papers she concludes that there were three non-empirical decision vectors working against Turin's theory: the

pressure), motivational (e.g., pride, conservativeness) and cognitive factors (heuristics) as well as theoretical values (e.g. elegance, simplicity).

conservativeness of the academic community, the protectiveness of the perfume industry and Turin's unprofessional behavior resulting in a lack of respect granted to him.

Only later did two non-empirical decision vectors emerge that shifted the distribution of decision vectors in favor of Turin's theory: the publication of his biography and the fact that he became the CTO of a company that successfully synthesized molecules according to his theory. For Solomon this adds up to a 3:2-relationship of non-empirical decision vectors working against Turin. In the case of this unequal distribution of non-empirical decision vectors dissent is *not* normatively appropriate. It were appropriate if the distribution was equal. Since Solomon seems convinced that Turin's theory is worth being pursued she therefore argues for manipulating the distribution of non-empirical decision vectors to become equal. And the easiest form to do this is to add a decision vector that pushes the distribution towards Turin. Solomon also has some suggestions of how this could happen: she imagines a BBC documentary, the publication of his book in the US as well as some public support of some prominent scientist. If one of those means succeeds, the goal would be fulfilled: the distribution of non-empirical decision vectors would be equal, dissent would become appropriate and the funds could be re-arranged to fund some of Turin's research.

To my mind, there are several problems with such an account of the distribution of cognitive labor, of dissent and diversity. First of all, as I have argued in detail in Part 2, the idea of simply *adding up decision vectors* seems to overly simplistic and misleading. Secondly, I am not sure whether Turin's role of a CTO that successfully produces molecules according to his theory is a *non*-empirical decision vector. After all, this success is an empirical support for his theory. Thirdly, it is left utterly untouched what would happen if there were two decision vectors in favor of Turin's theory (a BBC documentary and some big name supporting him). This would make the distribution unequal again and would imply that there should be no dissent anymore. Yet what instead? Consent on Turin's theory? This is clearly not very likely to happen just because Turin publishes his book in the US combined with a BBC documentary. I think that taking Solomon's examples serious, as I do here, shows the invalidity of a procedure of simply adding up unweighted decision vectors.

Moreover, it is also not quite clear why randomly chosen decision vectors are needed altogether to address the issue of diversity even within Solomon's own approach. After all, she rightly asserts that as long as more than one theory can account for empirical data that the mainstream approach cannot explain all those theories should be funded as well. An interesting question would refer to the exact distribution of funding, to the percentage of money that should be given to competing approaches. Yet for this crucial question Solomon does not give an answer. I agree this is a difficult, maybe even an impossible task. However, if it is not even tackled, it remains unclear why we would have to quantify decision vectors and manipulate them in the first place.

Further, I see epistemological - as well as ethical - problems in her recommendations of how to manipulate the non-empirical decision vectors. While I certainly do not question the importance and merits of scientists writing popular science books and engaging in TV documentaries, I wonder whether these activities should be decidedly recommended as means of encouraging epistemic diversity by changing the distribution of non-empirical decision vectors. After all, you can make all sorts of documentaries and write books of varying quality on various topics. Yet I do not see how this should raise the fundability of a particular theory of their authors. It would have to be assessed to what extent such maneuvers are positively or negatively affecting the distribution of research money over competing approaches. Any social epistemology aiming at informing science policy should do better than simply offering the recommendation to give money to those who make the most noise.

Finally, seeking alliances with big names to gain power may well be a successful strategy of changing the distribution of decision vectors to one's favor. Indeed, the relevance of alliances has been stressed in particular by many STS scholars. However, this acknowledgment of such strategic alliances has been attacked quite vehemently by several analytic social epistemologists. (cf. for instance Goldman's critique of Latour's "political-military account of science" (Goldman 2003: 225f). How then did such strategies and power games suddenly get rehabilitated? Sure enough, such strategies often prove helpful. Yet whether they are epistemically – and ethically – justified and should be recommended, is – or at least should be - a whole different story.

11.2.2 Kitcher's Laissez-Faire Position

Another author who has addressed the distribution of epistemic labor is Philip Kichter. I have portrayed his analyses in some detail in Chapter 6. The main question concerning the topics raised in this section reads as follows: "[W]hat is the optimal division of labor within a scientific field, and in what ways do personal epistemic and non-epistemic interests lead us toward or away from it?” (Kitcher 1990: 22).

To elucidate why Solomon considers Kitcher’s approach to be laissez-faire and to understand how and why it differs from invasive accounts, let me recapitulate his main claims. Kitcher develops different scenarios of how a scientific community may distribute itself if there were several rivaling methods or theories on a given topic. One example that he chooses is finding the structure of the DNA either with the help of X-ray cristallography (method 1) or by a second method involving “guesswork and the building of tinker-toy models” (Kitcher 1993: 11). He argues that before the DNA-structure has been discovered, everybody agreed that the likelihood to discover the structure of the DNA with the help of X-ray cristallography was assumed to be far greater than by using method 2. Hence, from the perspective of each individual scientist it seems most rational to pursue method 1, because this is the more promising theory. However, if every scientist acted like this, all scientists would use method 1 and none would use method 2. As it turned out, method 2 was successful in the end, thus it would have been detrimental for science if no one had pursued it. Kitcher argues that in this case, individual rationality and the community rationality are at odds: what is rational for the individual has negative consequences for the attainment of the community goal.

Kitcher offers in principle three solutions to solve the problem of suboptimal distribution of cognitive labor: a philosopher-monarch who assigns tasks to scientists and two modes of self-allocation: altruistic and egoistic. In the case of the philosopher-monarch, someone with complete control over scientific workforce could allocate scientists to different approaches, if he or she knew the likelihoods with which each approach would be successful. Since neither philosophical monarchy nor centrally planned economy has yet been successfully installed for science policy, Kitcher opts for self-allocation. Yet to make sure that a premature consensus on the most promising theory or method is avoided one may need some altruistic scientists to pursue the less promising approaches for the good of the community. Altruism, however, may not be needed to ensure a distribution of cognitive labor over different approaches and Kitcher asks how a Hobbesian community of completely self-interested egoists would distribute its labor. What if the person who discovered the structure of the DNA was promised an attractive price? Kitcher argues that “the probability of your winning is the probability that someone in your group wins, divided by the number of group members” (Kitcher

1990: 15). Hence, if *being first* is a goal, e.g. because one is offered a price, this motivation can lead to better distributions of labor than if everyone only strived for truth and hence chose the most-likely theory. This implies that *epistemically sullied agents*, credit-driven agents that rather aim at being first and being rewarded are actually beneficial for science as a whole. A scientist's quest for reputation and personal glory would be a necessary condition to avoid premature consensus in science and enable better distributions of cognitive labor. Kitcher concludes that "[s]ocial structures within the scientific community can work to the advantage of the community epistemic projects by exploiting the personal motives of individuals" (Kitcher 1990: 21)

Solomon interprets Kitcher's perspective as laissez-faire, because being credit-driven appears to be sufficient to enable an optimal distribution of cognitive labor. Hence, there is no need to interfere with science by exerting influence on the distribution of labor. As if an invisible hand was at work, scientists allocate themselves to different approaches in adequate ways by simply following their grubby motives. This assertive, "realist vision of science" has earned critique not only from feminist philosophers of science, but also from philosophers of science and social epistemologists more broadly. First of all, it has been noted that while there may be no philosopher-monarch, science funding indeed does have an impact on the distribution of labor, an impact that gets neglected in Kitcher's perspective. Moreover, the assumption that science works just fine or even as good as it possibly could has been disputed (Leplin 1994 1994: 666f, or Fuller 1994). Moreover, not only Solomon's reference to cases of premature consensus, but also feminist analyses concerning the detrimental effects of biases on the funding of non-mainstream research should give enough reason to the contention that a pure laissez-faire model may not suffice in enabling a sufficient amount of diversity in science (e.g. Harding 1991, Tuana 1989). Hence, diversity may have to be supported via science funding, non-mainstream research has to explicitly be encouraged and scientific pluralism should be promoted. Yet to what extent and how exactly funding should be distributed over different approaches is by no means clear.

11.3 Key Topics & Examples

11.3.1 ES^S & Epistemic Tasks: Tapping Diversity to Find & Create Knowledge

ES:Selection is crucially based on the acknowledgement that knowledge and skills are distributed over epistemic agents. To tap this distributed knowledge and skills it is necessary to attract many diversely equipped epistemic agents, which is a commonality with the other two types of epistemic sociality and ES^A in particular. However, ES^S differs from ES^A not only with respect to its form of closure, but also with respect to the epistemic tasks for which it is suited.

ES^S can be employed for two quite generic epistemic tasks: finding existing knowledge and creating new knowledge. Quite often the knowledge one needs exists somewhere – one simple does not know where. The best way to find it is to ask as many people as possible about it. And ICT and the Web in particular are well suited to enable the bi-directional communication to large numbers of people needed for this task. Two examples should suffice to illustrate the merits of ICT to find existing knowledge.

Example 1: My notebook started to make strange sounds few weeks ago. I assumed there must be something wrong with the fan, but I was not sure how to solve this problem. I posted a question on facebook.com, describing the problem and asking for advice. I not only knew that some of my facebook-contacts have the same computer, many of them were also computer scientists or quite technically literate. Sure enough, after a few hours I got several replies and a link to some software. I installed this free software and was able to reduce the noise. Then I reported about the benefits of this intervention on facebook as a feedback to those who offered help and as further evidence for those with potentially the same problem.

Example 2: On one of the philosophical mailing lists I have subscribed to people frequently pose questions, such as: Who coined the phrase "coal pit of induction"? Where did Foucault comment on others classifying him as a poststructuralist?[180] Often sources for quotes of references are asked for. An example is the inquiry below.

[180] These questions were asked and replied upon on the PHILOS-L@liverpool.ac.uk – mailing list.

> “Gilbert Ryle is said to have considered Zeno's paradoxes as representing the quintessential philosophical problem. Can anybody provide me with the reference to where he said this as a full bibliography reference? “

This question was asked on the mailing list PHILOS-L@liverpool.ac.uk on 15.01.2010. In three hours after the posting the inquirer reported back to the mailing list that the reference has been located and shared this reference. In other cases, people inquire for suggestions on good introductory literature on various topics. Usually, those who post the inquiry receive replies off the list. Afterwards they often thank their respondents and provide either the result or a summary of the replies to the list.

In both examples, an epistemic need existed: I needed knowledge about the possible sources of noise coming out of my computer. Someone else needed knowledge about the origin and source of a quote or recommendations about introductory literature. In both cases, this epistemic need could be satisfied by tapping the distributed knowledge of those agents that received the request and felt inclined to reply. Referring to a *tapping of the wisdom of the crowds* may seem appropriate to characterize this epistemic use of ICT. However, it should be noted that in Surowiecki’s usage of the term, wisdom of the crowd implies a form of aggregation (Surowiecki 2004). But for ES^S such aggregation is not needed. The epistemic content that is searched for already exists; it does not have to be created via aggregation. The knowledge is out there, all we need is to find it.

However, ES^S is not only at work when we want to find or distribute existing epistemic content. It can also be employed to *create new knowledge*. In this case it is not only the distribution of existing knowledge, but also the distribution of different skills that is crucial for the epistemic merits of ES^S. As noted before, the paradigmatic example for ES^S for the creation of knowledge is the distribution of effort within scientific communities and the role that science funding plays for this process. Within a specific scientific domain, a crucial goal of science funding is to distribute money over different approaches, because it is unknown which approach, which method, which theory will lead to the desired results. Or think about a patient group for diabetes. Assume that they want to use the money they have raised to fund research. They only have a certain amount of money to spend and several alternative approaches exist. Some of them are more promising than others. Should they donate all the money on the most promising approach? Or distribute it over several approaches? If so, what would be a recommended distribution? Questions like these have been discussed by several social epistemologists, as outlined before. A different solution to questions concerning the distribution of epistemic labor is offered by price-based mechanisms that rely on self-allocation and the most prominent example of a system employing such mechanisms for the creation of knowledge has been offered on the Web: Innocentive.com.

11.3.2 ES^S on the Web: The Case of InnoCentive

As noted in Chapter 2, InnoCentive describes itself as a "prime example of open innovation's crowd sourcing model."[181] It is based on the idea that the knowledge and the skills to solve so-far unsolved problems exist but are distributed over different, unknown agents. Hence, the epistemic benefit of a socio-technical epistemic system like InnoCentive lies in its ability to attract a large amount of epistemic agents who possess different skills and abilities. InnoCentive serves as a broker between those who have identified a problem and want it to be solved and those who possess the means to solve it. The epistemic goal of InnoCentive is to create new knowledge, solutions to pressing scientific, technical or medical problems. In this sense, InnoCentive is the example of epistemic social software portrayed in this thesis which in its mechanisms and goals is closest to science – and it is a prime example of ES^S. Recall the main characteristics of InnoCentive:[182] InnoCentive is a *price-based* system, which means that one or several solutions to a problem posed will be awarded with a price. To be awarded with this price the problem solution has to be *submitted in time* and the specified *criteria for success* have to be fulfilled. The *selection mechanism* involves the *problem seeker* as well as the InnoCentive staff as a mediator. They decide whether the criteria were satisfied. If several proposals were submitted, they decide which or how many submissions will receive the price or a share of it. The exact mechanism of selection differs between different challenges and is not outlined in detail. The criteria for selection also differ but are outlined in the challenge description.

The requirements are lowest and least specific for *Ideation Challenges.* Since these function primarily to brainstorm new ideas and require often only few written pages the barrier for participation and submission is low. Two examples introduced in Chapter 2 concerned finding ways to motivate people to use public transportation to reduce greenhouses gases in Chicago as well as ideas about how to improve banking processes

[181] www.Innocentive.com/_.../InnoCentive_Corporate_Overview.pdf [date of access: 15.02.2010]
[182] For a more detailed description please confer Chapter 2.

in the developing world. These questions are clearly crucial and as important, if not more important, than solving a more specific questions, such as finding a statistical approach for glucose monitoring. However, the questions are rather broad and meant to induce a brainstorming of ideas that do not have to be well developed. Often a 2-3 page description is sufficient for submissions to such *Ideation Challenges.* At the time of selection, the merits of the proposals cannot be based on a direct verification, which means that the selection criteria and the selection process are more subjective.

This is different for *Theoretical Challenges* and *Reduction to Practice Challenges*. *Theoretical Challenges* require more than just a short description of an idea. The examples provided in Chapter 2 were finding new methods to analyze consumer emotions and the above mentioned statistical approach to deal with the variability of data collected during continuous blood glucose monitoring. In the first example a detailed description of an experimental design is required for submission, in the second a statistical method has to be explained and a source code had to be provided. A *design* of the solution is required and since more effort and more skills are required, there are fewer submissions to be expected for *Theoretical Challenges* than for *Ideation Challenges*. However, the selection process becomes less subjective, because scientific standards for the statistical methods and experimental designs exist according to which the submissions can be assessed.

Finally, *Reduction to Practice Challenges* are the challenges for which most effort and most skills are needed. Here problem solvers need to provide *evidence*, i.e. empirical data or physical samples. They need to provide proof for their claims and neither the idea nor the design of a solution is enough. Instead the proposal needs to be validated by the solver and this validating must be replicable by the problem seeker. As a result, there will in all probability be even less submissions to *Reduction to Practice Challenges* than for other types of challenges, simply because more skills and efforts are required to submit a proposal. Moreover, the selection process is highly specified. The fewer the submissions and the clearer and more easily verifiable the success criteria are, the easier the selection process is. The validation in the case of *Reduction to Practice Challenges* comes closest to quality control mechanisms in the sciences. The difference is that on InnoCentive it is not the peers who assess the quality of a submission, but those who announced the challenge. Yet to make sure that they do not shirk paying the price money, this process is supervised by the InnoCentive staff.

By posting a challenge on InnoCentive and offering a price for those who can solve it problem seekers aim at tapping the distributed knowledge and skills of the more that 200.000 problem solvers from more than 200 countries which are registered at InnoCentive.[183] The exact solution to their challenges is probably not out there; at least it is not as easily trackable as the source of a quote or the solution to computer noise. Hence, InnoCentive is not about finding knowledge, but about creating knowledge. Nonetheless, existing knowledge and skills are needed to create this new knowledge. This existing knowledge and the skills are distributed and hence may be found in unknown epistemic agents which are part of the InnoCentive pool. The task is to find the right person to solve the problem, to find the needle in the haystack. The likelihood that a person with the knowledge and skills needed exists increases with the number of potential problem solvers and their diversity. And the likelihood that the problem will be solved depends on this size and diversity of the pool of problem solvers as well as on the incentives for tackling the challenge. The former issue depends upon the visibility of InnoCentive, the latter on the price structure.

In its selection mechanism, InnoCentive is a prime example of ES^S. One or several epistemic results are chosen to be awarded with a price. However ES^A and ES^I can play a role at different stages of the process. Not only can ES^I and ES^A-type mechanisms be employed to create the solution that is going to be submitted to InnoCentive. The selection mechanisms itself, the judgment about the respective merits of the submissions can be based upon ES^A, e.g. by *voting* on the proposals, or on ES^I, e.g. by *deliberating* on the proposals.

11.3.3 ES^S: Mechanisms & Criteria for Selection & Validation

ES^S is employed to find existing knowledge or to create new knowledge by making use of distributed knowledge and skills. Although the selection mechanisms can be based on mechanisms based on ES^I and ES^A, no method of re-combination is essentially needed for this type of epistemic sociality. ES^S is social, because if selects among the results that different epistemic agents provide for the same task. Selection still refers to a variety of different processes, such as the differences between dictatorial, consensual or voting-based selections. Taking a look at all the examples provided before it becomes

[183] http://www.Innocentive.com/crowd-sourcing-news/innocentive-at-a-glance/ [date of access: 29.03.2010]

obvious that not only the mechanisms of selection differ. The *criteria of selection*, the *criteria for validation or verification* of the selection also differ. The comparative merits of these *different mechanisms of and criteria for selection,* their respective pros and cons, the interaction between different types of epistemic sociality for selection, etc. should be of central interest for social epistemology.

While issues related to different integrative and aggregational *mechanisms* have been discussed in the previous two chapters, the question concerning the *criteria* for selection leads us back to the debates about shared norms and standards of evaluation (Longino 2002c), to the criteria for technical or empirical success (Solomon 2001), to the questions of how epistemic content may be ascribed the social status of being knowledge (Kusch 2002). I have positioned myself with those accounts which consider knowledge to be a social status and an honorific term, a qualifier which can be ascribed to some content and rejected for other content (Kusch 2002, Longino 2002c). In such a view, the creation of knowledge depends upon a community to ascribe the status of knowledge to epistemic content and this process implies deciding upon and selecting content for *quality*. However, what is meant by quality and how it can be assessed varies between different communities: for standards to be binding, they have to be shared within a community, but they may differ between communities. Accordingly, there are different views on what is meant by *quality* and how it can be assessed. Indeed, the topic of quality seems to have become a hot topic recently. One example is a conference on QualityCommons, which was held in Paris in January 2010.[184] On this two-day workshop, different positions on *quality* and what this term may mean in science, arts and on the Web were discussed. One of the questions was whether quality is an objective characteristic of objects or whether it always lies in the eye of the beholder. In line with what has been argued throughout this thesis, I consider quality to lie in the eyes of a community of evaluators who share standards and criteria. Hence quality assessment is relative and contingent to those communities – but this does not imply that it is completely arbitrary.

The relationship between the availability of shared standards and the ease by which quality assessments and selections can be made has also been documented by Michèle

[184] The workshop QualityCommons has been organized by the Centre for Research in Social Simulation (University of Surrey, UK) and the Centre d'Analyse et Mathématique Sociales (EHESS/CNRS, France). Further information can be found at http://cress.soc.surrey.ac.uk/web/events/qualitycommons-workshop/ [date of access 21.03.2010]

Lamont's comparison of peer review processes in the social science and humanities (Lamont 2009). Analyzing grant review panels in different disciplines she shows that while in some disciplines there seems to be a consensus about definitions of excellence, e.g. history and economics, other disciplines lack such consensus on criteria for excellence. One of the conclusions to be drawn from her analyses seems to be that the clearer the methodological standards and methods of a discipline are, the easier it is to reach consensus on the quality of epistemic content and epistemic agents and hence to select among them.

Taking a look the examples of ES^S used before, it becomes obvious that different criteria for selection and validation of proposed solutions have been employed. In the case of the inquiries for the sources of quotes verification is easiest. If available or accessible, one only needs to check the proposed source to verify the proposal. In the case of the computer problems, verification is also possible, albeit not as directly as is the case for the quotes. In comparing the noise level before and after the proposed intervention one can *infer* that the intervention caused this change and use this to draw conclusions about the validity of the intervention.

Things get already more difficult in the case of the best book on social epistemology, because there is no independent reality with which the selection can be compared to ensure its validity. Here, the selection is a matter of *comparative quality assessment* and not a matter of *match*. The same is true when people ask for good introductory books on philosophical mailing lists. There is no official list or ranking of books to which the results can be compared to. Yet the recommendations received can be more or less useful. At the very least they may have the epistemic benefit of offering a pre-selection of epistemic content and by doing this save the inquirer time to search for them. They function similar to recommender systems – it is the yet un-automated form of advice and this form of advice may exhibits the same epistemic merits and dangers as those outlined for RSs in the previous chapter. This lack of a direct means to assess the validity of recommendations refers to an inherent problem of quality assessment of epistemic content and I return to this topic below in a section on the selective function of peer review and ratings.

Finally, when it comes to creating new epistemic content, new knowledge, as is the case of science or on InnoCentive, different criteria for selection can be proposed. I have shown for InnoCentive that depending on the type of challenge the criteria for success are more or less specified imposing more or less restrictions on the selection process. The framework from social epistemology that seems to be best suited to understand the validation of the selection process seems to be Longino's notion of *conformation* (Longino 2002c: 116ff). As outlined in more detail in Chapter 5, c*onformation* is an umbrella term for the empirical success of epistemic content of which notions such as truth, isomorphism, homomorphism, fit, similarity or approximation are just special instantiations of. Crucially, conformation depends on the *purpose* of an epistemic task, it comes in degrees and it allows for epistemic pluralism.

11.4 Implications for the Analysis and Design of Socio-Technical Epistemic Systems

11.4.1 Epistemic Diversity: Laissez-Faire for the Web, Intervention in Science

An important question for the analysis and the design of socio-technical epistemic systems concerns the distributed epistemic processes that take place before their termination through selection and the role that diversity plays for them. Hence, one may ask what implications can be drawn from Kitcher's (Kitcher 1993) and Solomon's (Solomon 2006) analyses of socio-epistemic processes in science for the relationship between diversity and the distribution of epistemic labor on the Web. Although for the reasons outlined in the section on *ES^S and Social Epistemology*, I consider Kitcher's laissez-faire approach insufficient to ensure diversity in science, it seems much better suited to describe the relationship between diversity and the distribution of epistemic labor on the Web. The two points concern the difference between task attribution and self-allocation as well as the role of prices to motivate people.

As Benkler has noted, self-allocation is one of the main strengths of commons-based peer production (Benkler 2006) and the Web seems to be an environment in which systems that allow for self-allocation flourish, because they tap a large and already diverse pool of people. Hence, Kitcher's invisible hand model of pure self-allocation might work better in a Web environment than in science, simply because there are more people that can allocate themselves to tasks. Therefore, the likelihood that *someone* will end up pursuing an approach that will be successful is higher, because the sheer number of people involved can be higher. The challenge here lies in *attracting* a large enough number of people to work on the problem posed. Here, prices come into play. Although financial incentives may not play a great role for the more modular and granular tasks that Benkler refers to, financial incentives may play a role if bigger chunks of work need to be done, if less modular challenges have to be met. As we have seen in the case of InnoCentive, some of the problems for which solutions were sought required not only substantial knowledge and skills, but also considerable effort, effort which cannot be as easily distributed as writing a Web encyclopedia or a source code. Also, financial incentives may play a role, if solutions to a problem are for commercial organizations. InnoCentive hence seems to perfectly exploit ES^S on the Web by harvesting a large number of diversely skilled epistemic agents and by combining financial incentives with self-allocation to encourage them to tackle the problems posed.

To my mind, a laissez-faire-approach, such as the one proposed by Kitcher may be suited for socio-epistemic practices on the Web, but less so in science. Given the potential epistemic agents that can be addressed on the Web, there is enough diversity. One only needs to tap it by being visible and combining self-allocation of tasks with a price structure that serves as an incentive for participation. Hence, tapping ES^S on the Web depends less on actively increasing diversity, but more on finding the means to tap it. In science by contrast, and within certain scientific fields in particular, diversity should be actively supported to avoid biased research funding and premature consensus. To my knowledge no explicit model has been proposed to give advice on how exactly diversity should be supported via science funding. Although Solomon's *Social Empiricism* is meant to deliver such a model it fails to do so for the reasons outlined before. Hence, for the moment we may have to settle for the acknowledgement that epistemic diversity and scientific pluralism should be supported by funding approaches that are perceived to have lower chances to succeed. Yet the percentages, the exact distribution of research funds may be impossible to assess beforehand.

Finally, the likelihood that someone will pursue the ultimately successful approach is not the only reason to support diversity in science or on the Web. Although Kitcher (Kitcher 1990: 11) presents the example of the x-ray cristallography and the method of tinkering and guessing as a choice between two methods, the discovery of the DNA structure was rather a result of a *combination* of different methods. The discovery of the DNA was the result of differently skilled, diverse people *collaborating* on solving a problem. It was an interdisciplinary team of a zoologist, two physicists and a chemist, trained in x-ray cristallography that solved this puzzle.[185] Hence, the "true story" of the DNA discovery is rather an example of ES:Integration, although Kitcher's abstraction seems to imply that it is a case of ES:Selection.

To conclude, diversity seems to be as beneficial for ES^S as it has been for ES^I and ES^A: when it comes to problem solving, the more diverse seems to be the better. However, there are some differences between the notions of diversity employed in different types of epistemic sociality. In the case of systems employing aggregational mechanisms, diversity was needed to avoid *statistical bias*. Diversity in ES^I and ES^S is not reduced to such a statistical necessity. Yet while in the case of ES^I, diversity was addressed on the level of research teams, diversity in ES^S is crucial on the level of communities as a whole. Finally, for ES^I, diversity has limits, because epistemic agents need to communicate with each other. For ES^S this common ground is much less needed.

11.4.2 Peer Review: A Case of Nested ES – Using ES^A and ES^I for ES^S

In the previous chapter I have shown that on the Web, rating and ranking mechanisms often serve to filter content. In science, there is one classic mechanism of filtering and quality control: peer review. It should be noted that peer review serves epistemic purposes other than just *selecting* content. Its quality-enhancing relevance also lies in *error detection* (e.g. in identifying error in theories, data or methods), *readability improvement* (e.g. tips to improve language, identification of grammar and spelling mistakes, identification of missing or inaccurate references, etc), making suggestions for *future research,* as well as the detection of *scientific fraud*, etc. However, given the topic of this chapter, in the following I focus on the characteristics and relevance of peer review as a mechanism to select high-quality content only. From this perspective the following questions are of interest:

a) How do peer review processes function as a socio-epistemological process of selection?
b) How can peer review in science be grasped with the help of my tripartite model of epistemic sociality?

[185] Please confer for instance: http://www.chemheritage.org/classroom/chemach/pharmaceuticals/watson-crick.html [date of access: 15.12.2009].

c) How does peer review differ from rating and ranking mechanisms on the Web?
d) Comparing different socio-epistemic selection processes, which ones are better suited to select for quality?
e) Could peer review as currently practiced in academia be improved by employing forms of open rating and ranking similar to those employed on the Web?

These questions are hard to tackle and even harder to answer. Hence, in the brevity of this section I cannot answer them to any satisfying extent. Instead I outline issues that need to be addressed if these questions should be tackled as well as problems that need to be solved to analyze the respective merits of different socio-epistemic processes of selecting content for quality control.

The first problem that arises when attempting to answer the questions posed above is that peer review refers to a wide variety of different practices.[186] First of all, peer review can be *open, blind* or *double-blind*, depending on whether the reviewer and/or the reviewed are aware about who is taking part in the process. It can be done with respect to *different objects*: epistemic content (e.g. journal papers, conference proceedings submission, grant proposals) or to epistemic agents (e.g. scholarships, fellowships, awards) can be assessed. Further, *different selection criteria* are being employed, which can be more or less explicit. And finally, *different mechanisms of judgment formation* can be employed: varying numbers of peers can assess the merits of epistemic content or epistemic agents via *rating* them or via *deliberating* on them, combinations are of course possible.

From the perspective of this thesis, these two forms of collective judgment formation refer to the difference between aggregational and integrative mechanisms: in other words, rating mechanisms in peer review exploit ES^A for ES^S, deliberative mechanisms exploit ES^I for ES^S. ES^A for ES^S is at work, when human epistemic agents rank or rate other epistemic agents or the epistemic content they have provided with the goal to select some, but not others. ES^I for ES^S is at work, if human epistemic agents deliberate about the merits of different other epistemic agents or the epistemic content they have provided to select some, but not others. Due to this nesting of different types of

[186] For a review on peer review in academic please confer the state of the art report of the EU-project LiquidPub: available at https://dev.liquidpub.org/svn/liquidpub/papers/deliverables/LP_D1.1.pdf [date of access: 30.03.2010].

epistemic sociality, various issues raised in this chapter as well as the two previous ones may be more or less relevant to scrutinize different peer review mechanism.

ES^I for ES^S

Michèle Lamont's before-mentioned analyses of peer review processes involved in several national funding competitions provide some examples of ES^I involved in ES^S (Lamont 2009 :2). In these competitions, multidisciplinary panels decided about the distribution of fellowships and grants in the social science and humanities. To analyze what she calls *evaluative cultures* and how they may differ between disciplines, she observed the deliberations taking place in the panel meetings and conducted interviews with the panelists. Two key insights that can be obtained by her analyses concern the differences and similarities between different disciplines. Lamont shows that while disciplines differ with respect to the clarity and bindingness of criteria for excellence, they share rules of deliberation, which facilitate consensus formation (Lamont 2009: 6). Lamont also argues that for grant review, as opposed to reviews for selecting conference or journal submissions, face-to face meetings and deliberation are favoured over quantitative techniques, such as citation counts. These meetings then are characterized by ES^I-type socio-epistemic processes and that all those issues on trust, authority, shared norms and values raised in Chapter 9 are relevant to understand and evaluate the socio-epistemic processes involved in this type of peer review can be documented by the following quote:

> "Debating plays a crucial role in creating trust: fair decisions emerge from a dialogue among various types of experts, a dialogue that leaves room for discretion, uncertainty, and the weighing of a range of factors and competing forms of excellence. It also leaves room for flexibility and for groups to develop their own shared sense of what defines excellence—that is, their own group style, including speech norms and implicit group boundaries. Personal authority does not necessarily corrupt the process: it is constructed by the group as a medium for expertise and as a ground for trust in the quality of decisions made" (Lamont 2009: 7).

Hence, it seems that not only deliberative peer review processes can be conceived as examples of ES^I employed for ES^S, but also that the framework I have proposed is well suited to point to important issues that deserve critical scrutiny in these processes. Hence, while Lamont asserts that in her case studies authority did not corrupt the process of consensual decision making, she thereby also indicates that whether or not authority corrupts consensual decision making is a valid and important question for analyses of peer review processes in general. It has to be analyzed whether fair decisions emerge from dialogue or not, whether weighting of criteria as well as differential weights given to the judgments of different participants have been conducive for selecting quality or not, etc. And although Lamont claims that "deliberation is viewed as a better tool for detecting quality than quantitative techniques such as citation counts", it is indeed a matter of further empirical assessment to show whether and under what conditions this perception is justified.

ES^A for ES^S

Citation counts are one quantitative way of approximating quality without peer review. However, there are also peer review processes that quantify their judgment formation by using rating and ranking mechanisms. As opposed to the deliberative processes described above, these peer review processes exploit ES^A for ES^S. In the following I focus on such rating and ranking-based peer review processes, but before, I need to address a question that is crucial for both deliberative and aggregational procedures: the concept of *quality* itself. I am addressing this topic at this point, because it is central to understand the problems that occur when trying to assess the comparative merits of peer review and although this problem becomes more lucid when trying to assess peer review quantitatively, it is as relevant for qualitative analysis.

A Note on Quality

One of the crucial goals of peer review is selecting the *best* content, selecting the *best* agents. This implies an assessment of quality. However, as argued before, quality is by no means a clear concept. Quite to the contrary, while its relevance for selection is uncontested, its meaning, whether it is an objective characteristic of objects or lies in the eye of the beholder is a hotly debated topic not only in science, but also in art and on the Web. One aspect of this debate that is of crucial relevance on the Web, but not only on the Web, refers to the relationship between quality and popularity. What is the relationship between the two of them? Can popularity be an indicator of quality or is popularity the same as quality? Are popularity and quality completely unrelated, as objectivist understandings of quality could imply? Topics like these ones have been addressed, but not answered at a workshop on Quality Commons, held is Paris in

January 2010.[187] Different positions can be taken and I consider workshops like the one mentioned to be an indicator of the pressing need to assess questions concerning quality. My stance towards quality is related to my socio-epistemological position more generally. I consider quality to neither reside in the objects to be reviewed nor in the individual subjects, but to lie in the eyes of a community of evaluators who share certain standards and criteria. This implies that quality assessment fundamentally depends on those communities and may differ between communities. It does not imply, however, that quality is completely arbitrary, because its assessment is based on such shared standards and criteria, which may change over time and be subject to scrutiny themselves. Nonetheless, they are temporarily binding. This understanding of quality as a community-based qualifier is based on Longino's concept of shared standards for the assessment of knowledge claims in science (Longino 2002c) and related to Kusch's analyses on the performativity involved in epistemic practices (Kusch 2002).

Keeping these difficulties in specifying quality in mind we can take a look at empirical studies trying to assess the merits of different forms of peer review. A crucial task with which such studies are confronted with consists in operationalizing an external *indicator of quality* with which the results of peer review processes can be compared. If we want to assess whether peer review can filter quality, can distinguish quality from junk, we have to define and operationalize quality in the first place.

Different studies on peer review use different indicators of quality with which they compare the review results. And unsurprisingly, they come to different results concerning the overall merits of peer review. For instance in an analyses of peer review in the case of three conferences in computer science, Casati and his colleagues compare the ranking of conference submissions with the number of citations that these papers received in the six years following their publication in the conference proceedings (Casati, Marchese et al. 2009). Hence, they used citation counts as the external, quantitative indicator of quality, as the criteria for validating the review process.

They find that with respect to this measure, the review process performs very poorly. Indeed they write that "the review process and the random selection process have the

[187] The workshop QualityCommons has been organized by the Centre for Research in Social Simulation (University of Surrey, UK) and the Centre d'Analyse et Mathématique Sociales (EHESS/CNRS, France). Further information can be found at http://cress.soc.surrey.ac.uk/web/events/qualitycommons-workshop/ [date of access 21.03.2010]

same quality in approximating the citation-based ranking" (Casati, Marchese et al. 2009). Moreover, they have analyzed the extent to which the selection process was influenced by various types of *biases.* Since the review process was double blind, certain types of biases could be avoided, such as affiliation, country or gender bias.[188] However, the selection was influenced by *rating bias*. Rating bias, as conceived in their analyses refers to the reviewers' tendency to either give consistently high or consistently low marks. More specifically, "[r]eviewers are positively biased if they *consistently* give higher marks than their colleagues who are reviewing the same proposal. The same definition applies for the opposite case, when we talk about negatively biased reviewers" (Casati, Marchese et al. 2009). The authors conclude their analyses by stating that in their analyses they found more randomness in peer review processes than presumed, that rating biases affect the selection process and that peer review needs to be improved overall. In specific, they propose to amend peer review with additional quality control mechanisms that are more open, transparent and efficient.

A different project by Marcel Weber and his colleagues analyzes the merits of peer review for science funding in biology and medicine.[189] In their analyses on the decision of grant funding conducted by the Swiss National Science Foundation (SNSF), the researchers come to quite different conclusions concerning the reliability, validity and fairness of peer review. They find no major biases, a high level on agreement on funding decisions between reviewers as well as "evidence that the decisions of a public funding organization for basic project-based research are in line with the future publication success of applicants"(Reinhart 2009: 789).

To assess the validity of peer review, the researchers compare the number of publications (in the four years after the granting decision 1999-2002) and citations (in the subsequent four years 2003-2006) that the main investigator achieved between projects that were funded and those which were not funded. The results they obtained by this comparison indicate "that average numbers for all included variables are significantly different between rejected and accepted applicants, confirming that those researchers who are successful in applying for grants from the SNSF continue to publish

[188] I ignore for the moment the fact that reviewers may be able to identify the authors of proposals even in blinded reviews as several studies seem to indicate, e.g. (McNutt, Evans et al. 1990), (van Rooyen, Godlee, et al. 1998).

[189] Swiss National Science Foundation: Quality Assessment, Expertise and Decision-Making in Scientific Research: Criteria, Procedures, and Social Organization, SNSF- funded SNSF-Professorship: PP001–102675.

more successfully than those who applied unsuccessfully. The authors conclude that based on these results the review procedure seems valid. However, this difference may simple be the result of being funded or not as the researchers acknowledge. Indeed, their further analyses showed that there are "not many significant correlations between assessments during the review procedure and the ensuing publication success" (Reinhart 2009: 805). Here, an interesting methodological aspect of this study becomes relevant. Besides recommending a rejection or acceptance of the proposal, the reviewers were not obliged to grade the applications numerically. Hence, the assessment of quality beyond the recommendation of acceptance or rejection had to be inferred, the textual recommendations had to be quantified. The researchers did this by inductively developing a code book with 19 quality criteria supplemented with a 5-point ranking scale (++,+,0,-,--).[190] Hence, a qualitative process was retrospectively transformed into a ESA-type process to be analyzed quantitatively.

What conclusions can be drawn from these two studies? First of all, different studies on peer review come to very different results concerning the validity of peer review as a means to select and filter quality. This difference can be partly explained by the fact that peer review refers not only to different processes (blind, double-blind, open, etc), but also can be applied to different selection processes. Selecting papers for a conference has different implications than selecting project proposals. The validity of the acceptance versus rejection decision cannot be assessed by comparing it with the citation counts. In the case of the conference, rejection makes citation impossible. In the case of the project funding, rejection makes publications of the epistemic agents less likely, because they do not have the funds to even conduct their research. This problem sheds some light on the more general problem of finding relevant proxies for quality with which peer review results can be compared to assess their validity. Moreover, how should peer review processes be compared if different proxies are used? Since there is no definite answer of what the quality of scientific papers amounts to, this quality has to be approximated by various measures, all of which come with their own shortcomings.

This leads us to the final question concerning the comparative merits of peer review versus web-based mechanisms to select and filter quality content. An important socio-epistemological question would be whether academic peer review systems work better

[190] This information was taken from a summary of a research report for the project that I received from the project leader Marcel Weber.

or worse in selecting high-quality content than web-based mechanisms such as the ones proposed in the previous chapter. To answer this question, we first need to understand how academic and web-based mechanisms of quality control differ. Once this analysis is done, we could also ask whether different traditional peer review could be combined with web-based mechanisms of post-publication control and if so, how exactly such a combination could look like.

From what was outlined before, it seems plausible to start by comparing selecting mechanisms that employ similar mechanisms, namely ratings and rankings. Rating and ranking systems on the Web serve as mechanism to filter epistemic content as well as epistemic agents. As such, they have a similar function as peer review has in academia. It is the major mechanism to filter epistemic content in scholarly communication for the assessment of grant proposals or as well as to filter epistemic agents when it comes to scholarships or fellowships. Hence it should be of socio-epistemic interest to compare the respective merits of peer review and different types of weighted (i.e. reputation-based) and unweighted (e.g. Amazon ratings) ratings systems. However, comparing these processes is not an easy task and so far there seem to be no conclusive results concerning the comparative merits of these different mechanisms.[191] There are several reasons for this. One cluster of problems is related to the difficulty of defining and operationalizing quality. Since quality can only be approximated by using different indicators of quality, comparisons will always depend on such operationalizations. Other problems concern the differences between web-based ratings and ratings in peer review processes more specifically. For instance, open ratings on the Web, such as Amazon ratings, and ratings on conference submissions differs with respect to the *distributions of rankings* as well as with respect to *exhaustion*.

The distributions of product ratings, such as the ratings on Amazon.com often have a distinct J-shape, i.e. many high marks (five stars), some low marks (one star) and very few ratings in between (Hu, Pavlou et al. 2009). The explanation for this distribution is that user tend to rate only products that they particularly liked or disliked, while mediocre products seems less worthy to be rated. Ratings of conference submissions do not have this characteristic distribution, they often have more rankings in the middle as the figure below visualizes exemplarily.

[191] However, work in this direction is currently being conducted within the research project LiquidPub.

Figure 21: Mark Distribution for the Overall Evaluation Criterion in a Conference[192]

Moreover, due to different allocation processes, *exhaustion* is another factor according to which ratings in peer review and ratings on the Web often differ. While in peer review processes for journals, proceedings or grant proposals, each submission is evaluated, on the Web only a certain percentage of items get evaluated. A crucial question that needs to be answered in order to compare or combine the selective power of web-based ratings with aggregational peer review therefore concerns the assessment of the relevance of these differences in distribution and exhaustion. What are the epistemic benefits and dangers of different distributions? How to account for the lack of exhaustion in open rating systems? With all these insights in mind, a simple answer on how to compare and how exactly to combine different selection mechanisms seems futile. Rather, more research is needed on the respective merits and shortcomings of different selection mechanisms on the Web and beyond. One of the crucial challenges of employing aggregational mechanisms for selection resides in tackling the relationship between *quality* and *popularity* – a difference that is particularly crucial for the assessment of *epistemic* content.

11.5 Conclusions on ES:Selection

In this chapter I have outlined my third and final type of epistemic sociality: ES:Selection. I have shown how it differed from the other two types of epistemic sociality, ES^A and ES^I as a distinct mechanism of closing socio-epistemic processes.

192 Taken from Casati, Marchese et al. 2009.

Instead of integrating or aggregating content provided by different epistemic agents, ES^S selects content or agents. ES^S therefore shares with ES^A the characteristic that the closure of the socio-epistemic process is done by agents other than the ones involved in the process itself: those who select are not the ones who are being selected or whose content is being chosen. Those who aggregate are usually not the ones whose content is being aggregated. Integration by contrast is often done by those who provide the content to be integrated themselves. This puts the selectors and the aggregators into a very powerful position. I have outlined the implications for ES^A in the last chapter by demanding close scrutiny of the aggregational mechanisms. The same is true for the selection mechanisms: the process of collective judgment formation upon which the selection is based should therefore be the primary locus of scrutiny in systems that employ ES^S.

In this chapter, I have also addressed the possibility of nesting different types of epistemic sociality for the first time. I have shown that ES^S often employs ES^I or ES^A at different stages within the socio-epistemological processes. Of course, it is also possible that ES^A and ES^I can be combined. This possibility of nesting is only more common and more visible it systems serving selective purpose. Indeed, the majority of ES^S-type systems employ ES^A or ES^I. Just think about the different peer review models which are either based on ratings or on deliberations. This nesting implies that whenever ES^S employs ES^I or ES^A, those issues addressed in the previous chapters become relevant. In the following table I therefore only add additional aspects that should be taken into consideration when analyzing or developing socio-technical epistemic systems that exploit ES:Selection.

Setting the Parameters for ES:Selection	
Nesting Types of Epistemic Sociality	• Be aware of the possibility of nesting different types of ES: Within socio-technical epistemic systems different types of epistemic sociality can be employed. This is particularly visible and prevalent in the case of ES^S, but exists for ES^I and ES^A as well. • If types of E^S are nested, their respective role as well as their interplay has to be analysed.
Diversity	• Tap distributed knowledge and skills on the Web. A crucial requirement to tap this diversity concerns the visibility of the web platform as well as adequate means to motivate epistemic agents to participate. For tasks which come in bigger chunks or require profound knowledge, skills and effort, financial incentives may be needed. • Actively induce and support cognitive and epistemic diversity in science: Cognitive diversity can be enhanced by interdisciplinary exchanges, by encounters with people who use different perspectives, heuristics, interpretations and predictive models (Page 2007). *Epistemic* diversity concerns the pursuance of different methods, approaches and theories. Here science funding plays a role in funding non-mainstream research.
Quality & Transparency	• Selection for quality depends on defining and operationalizing quality. According to the perspective adopted in this thesis, quality neither lies in the content or agents to be assessed, nor in the evaluator only. Rather quality resides on the *eye of the community* which discursively enacts and performs – or challenges and changes – shared standards of evaluation. • Selection processes differ with respect to the criteria and mechanisms employ. • The clearer the standards and the criteria, the easier it is to reach consensus in deliberative processes of ES^S. The transparency of standards and criteria makes it possible to refer to them, to address them – and to change them if necessary. • As was the case for the other types of epistemic sociality, transparency is a crucial requirement for socio-technical epistemic systems. Transparency here refers to the visibility of mechanisms, criteria and standards for selection.

Table 7 Setting the Parameters for ES:Selection

12 Conclusions, Limits & Future Research

In this thesis I have outlined a socio-epistemological framework to analyze epistemic social software applications, i.e. social software with a decided epistemic function. The reason for developing a new model was that although social epistemology appeared to provide an appropriate frame of reference for the analysis of such systems, none of the comprehensive social epistemologies provided so far appears to be sufficient. Clearly, none of them had the intention of being able to analyze epistemic social software and most of them were explicitly designed to analyze socio-epistemic practices and processes in science (e.g. Solomon 2001, Longino 2002c). Nonetheless, if one considers epistemic social software to have induced an important change in our epistemic environments, then it should be possible to analyze them from an epistemological perspective. And since the key characteristic of these systems concerns the inclusion and orchestration of multiple epistemic agents, a *socio*-epistemological analysis should be most appropriate.

12.1 Positioning within Social Epistemology

Throughout this thesis I have therefore tried to make insights from various social epistemologies fruitful for such an analysis. Almost all of the social epistemologies portrayed did find their way into my own framework. The main reason for this is that while none of them seems sufficient for understanding all socio-epistemic processes in socio-technical epistemic systems, all social epistemologies highlight crucial issues related to different *aspects* of such socio-technical epistemic systems and practices. Hence, my model is to a certain degree eclectic in making use of insights from different social epistemologies, which may well contradict each other with respect to certain aspects. Yet, despite this openness towards insights from all these approaches in order to understand different aspects of epistemic social software or socio-technical epistemic systems more broadly conceived, I have critically assessed and compared different social epistemologies with respect to the adequacy and fruitfulness with which they can apprehend the social nature of knowledge.

I have been particularly critical of approaches in social epistemology which neglect processes of *knowledge creation* in favor of processes of mere *transmission or distribution*. This focus on knowledge transmission seems to be related to certain narrow conceptions of *testimony* in epistemology. I agree with Kusch that any comprehensive account of testimony should account for the *performative* aspects of testimony and hence for its *knowledge-creative* function (Kusch 2002). Nonetheless, narrow conceptions of testimony as mere distribution of knowledge may be helpful to understand distributive practices on the Web.

Moreover, I have outlined the distortive dangers that reside in unreflective use of *quantification* in several social epistemologies. Based on the insight that a lot of trust is placed in numbers in our current society (Porter 1995), I have argued that we have to be very careful when quantifying the qualitative. For this reason I have criticized especially Goldman's application of Bayesian inference to testimonial evidence (Goldman 2003) as well as Solomon's counting of decision vectors (Solomon 2001). Quantifying the qualitative always entails the risk of bias (Friedman and Nissenbaum 1997). This risk does not imply that quantification is useless or should be avoided. It does imply however, that we have to be very careful in our processes of quantification, and more specifically, that we have to lay open these processes of quantification, its limits and the assumptions that they relied upon. In the case of Goldman and Solomon, quantification runs the risk of hiding the highly subjective nature of the information fed into the quantitative process, of casting a patina of scientific integrity over possibly quite faulty assumptions. How dangerous this can be has been also addressed extensively in my assessment of different types of epistemic social software.

I have argued that approaches which consider knowledge to be a social status that crucially depends upon communities, upon communication and critical interaction among epistemic agents are to be preferred. In particular, I have sided here with the approaches brought forward by Helen Longino (Longino 2002c) and Martin Kusch (Kusch 2002), being aware that there are crucial differences between the two models they propose.

A further aspect which I share with both Kusch and Longino concerns the receptiveness towards historical and sociological analyses of knowledge and science. While I agree that social epistemologies should not remain purely descriptive, any normative social epistemology has to take insights from empirical fields into account, because to my mind *normative appropriateness depends upon empirical adequacy.* Only once we understand socio-technical epistemic systems and practices we can start thinking about amending or improving them.

Concerning the issue of *complementarity*, I consider social epistemology not to be an addendum to individualistic epistemologies. Agreeing that knowledge is a social status implies that individuals cannot know and that hence, that knowledge only comes into existence though communal vetting of epistemic content. In this sense, social epistemology is not complementary to individual epistemologies, but a successor discipline. Nonetheless, while singular epistemic agents may not be able to *know*, they can surely *believe, see, infer*, etc. And for those processes that proceed or succeed these communal knowledge creation insights from individualistic epistemologies may be as much of use as insights from psychology or cognitive science.

Understanding knowledge to be a social status implies that knowledge fundamentally depends upon communities. I have further agreed that epistemic standards differ between communities. Hence, if the attribution of knowledge depends on such shared local standards, knowledge is always relative to these standards and the communities in which they are being enacted. I have accordingly sided with Kusch's relativist position (Kusch 2007, Kusch 2002). According to him *relativism* denotes that truth and falsity depend on a community's interests, exemplars and goals. It does not imply however, that all statements are only "relatively true" or that one can make a theory true by simply wanting it to be true. Relativity here means relative to standards and goals, not complete arbitrariness.

I have emphasized that *trust* is fundamental for all epistemic practices and that every socio-epistemological theory has to account for this fundamental entanglement of knowledge and trust and for its epistemological – and ethical – consequences. Acknowledging this entanglement does not imply that trusting is always justified. Instead of advocating blind trust, the notion of *epistemic vigilance*, as introduced by Sperber and his colleagues, seems to be more adequate to describe the relationship between knowing and trusting: we need to trust others to know, but humans seem to "have a suite of cognitive mechanisms for epistemic vigilance, targeted at the risk of being misinformed by others" (Sperber, Clément et al. to appear). To understand these mechanisms is of vital interest not only for the analysis of epistemic practices on the Web, but for any social epistemology.

A crucial issue that has been addressed throughout this thesis concerns the *distribution of epistemic labor.* This distribution can take place on different levels. To use a comparison with science: Distribution of epistemic labor takes place between two people, within research groups as well as within overall scientific fields. My socio-epistemological framework is a framework to understand different forms of such distribution through the way in which such distributed labor is *terminated.* My model proposes a tripartite classification of different mechanisms of closure terminating socio-epistemic processes in which multiple epistemic agents fulfill epistemic tasks and reach epistemic goals in a distributed fashion. In social epistemology, closure has primarily been analyzed in the context of theories on c*onsensus formation*. I have agreed with several social epistemologists that consensus is not intrinsically valuable or functions as the ultimate goal of science, and that epistemic pluralisms and diversity to avoid premature consensus should rather be advocated (Solomon 2001, Longino 2002c). However, even if *consensus* it not needed, *closure* of epistemic practices is essential for the creation of epistemic products. This is the difference between analyses on consensus formation and my account on mechanisms of closure: mechanisms of closure as conceived in this thesis refer to the process of temporarily terminating distributed labor to create an epistemic product which can then be shared. This closure can be achieved via the three different mechanisms I have proposed as the basis of my socio-epistemological framework: *integration, aggregation and selection*. Such mechanisms of closure function on a different level than consensus as conceived in most socio-epistemological accounts. Closure is the process by which epistemic products get finalized. It is neither the same nor does it depend upon macro-level consensus on scientific theories or methods.

12.2 A Social Epistemology for Socio-Technical Epistemic Systems

So far there have been only few attempts to analyze social software from a socio-epistemological perspective. While they offer some interesting insights, they can at best serve starting points for a comprehensive analysis of such socio-technical epistemic systems for several reasons. First of all, most analyses have focused on the *distributive* functions of ICT, while neglecting *creative* processes. With respect to social software more specifically, only few examples have been analyzed, Wikipedia being the most extensively studied system. Moreover, given the limited number of analyses, only few socio-epistemic aspects of epistemic social software have so far been addressed. My model is therefore an attempt to widen the perspective, to *expand the frames* (Suchman 2007/2009) for socio-epistemological analyses of epistemic social software.

This reference to Lucy Suchman's notion of *expanding frames* leads me to a more fundamental critique of previous socio-epistemological analyses of ICT: their inadequate models of technology. I have argued that social epistemology has not yet addressed the entanglement of the social, the technical and the epistemic to a satisfying degree and should therefore rely on insights from the field of Science and Technology Studies (STS) to shed some light on this relationship. More precisely, I have criticized the sharp distinction between human agents and technologies that often seems inherent in socio-epistemological analyses of socio-technical epistemic systems. Technology often seems to be conceived as completely predictable and controllable tool in the hands of fully rational human agents. This perspective on technology seems inappropriate to understand socio-technical epistemic processes within epistemic social software. Instead, epistemic social software should better be understood as socio-technical epistemic systems in which multiple human and non-human epistemic agents interact in creating epistemic content. I have relied on insights from numerous major accounts in STS, feminist STS as well as the field of *Values in Design* and the main insights that found their way into my own socio-epistemological framework concerned the acknowledgement of the *entangled nature of the social, the technical and the epistemic*; an assessment of the *differences and similarities between human and non-human actors*, a focus on the *role of values in technology design,* and the adoption of a *performative conception of socio-epistemic practices and systems.* My own model is therefore informed not only by social epistemology, but also quite crucially by my readings in the field of STS and its neighbouring disciplines.

Any socio-epistemological framework needs to set some premises with respect to the fundamental epistemic concepts it employs. I have adopted *Longino's tripartite notion of knowledge* as well as her concept of *conformation* as a criterion for the validation of epistemic content as a basis for my own model (Longino 2002c). I consider her differentiation into *knowledge as content* from knowledge as a set of *knowledge-productive practices* and *knowledge as cognitive agency*, i.e. as a state of a person ('knowing') to be a useful basis for analyzing socio-epistemic practices in socio-technical epistemic systems. More specifically, I have argued that *knowledge as content* can serve as *input and output* of *knowledge-productive practices* in which situated epistemic agents are involved. Longino's notion of *conformation* appeared also to be the most useful framework for analyzing socio-epistemic practices and products on the Web for several reasons: conformation can be applied to non-propositional content and to epistemic practices, it comes in degrees and it can be assessed for different dimensions depending on the epistemic goals.

12.3 Key Insights

As a result of my analyses and readings, I have proposed a socio-epistemological framework that is based on a classification of three important types of epistemic sociality. These types of epistemic sociality are distinguished and named after the mechanisms they employ to close socio-epistemic processes which are distributed over multiple agents: *ES:Integration* (ES^I), *ES:Aggregation* (ES^A), *ES:Selection* (ES^S). In systems employing *ES:Integration*, the singular epistemic results accomplished by different epistemic agents need to be *integrated* to achieve the overall epistemic goal. In systems employing *ES:Aggregation*, these individual contributions need to be aggregated. Finally, in systems employing *ES:Selection* neither aggregation nor integration needs to take place. Instead some epistemic agents, usually none of those who have provided the singular results, choose among those results. This process can be doubly social, if it is based on aggregational or integrative mechanism. Yet even in cases of dictatorial decision making, this process is socio-epistemic in exploiting socially distributed knowledge and skills.

This classification is not meant to imply that the closure mechanism is all that matters when it comes to assessing socio-technical epistemic systems. This is clearly not true and my analyses of the different socio-epistemic processes and stages throughout this thesis would clearly belie such an assumption. However, this classification functions as an *analytical tool* which not only sheds light on important differences between systems, it also highlights which issues are of particular relevance for the socio-epistemological analyses of such socio-technical epistemic systems. First of all, these three types of epistemic sociality differ with respect to their prerequisites, their strengths and weaknesses. Secondly, they differ with respect to several central topics of socio-epistemological interest which I have outlined for each of the types of epistemic sociality. While some topics are relevant for each type, e.g. *diversity* or *trust,* their relevance, form and loci differ between the three types of epistemic sociality. For instance, while diversity plays a role in all three types of epistemic sociality, the rationale behind this request for diversity differs between them. While for ES^I, diversity needs to be coupled with shared standards to make the integration of results possible, no shared standards are needed for ES^A. Here diversity refers to statistical variance which is needed to avoid systematic bias in the aggregation process.

Moreover, I have argued that *trust* is crucial in all socio-technical epistemic systems, irrespective of which mechanism of closure they employ. Nonetheless, the loci of trust differ between such systems. In selective and especially in aggregative systems, there is a concentration of trust on the aggregation and selection mechanisms themselves. This trust in the aggregational mechanisms has been addressed in the discourse around *algorithmic authority* (Shirky 2009). I have argued that algorithmic authority is a specification of *procedural trust*, of trusting a procedure, a mechanism instead of a person or institution. This form of trust is not new. Trust in different methods in science - be it in specific scientific methods or in general methods of quality control such as double-blind peer review - is also a form of procedural trust. The novelty lies in the fact that these mechanisms are often unknown to the users of systems which puts them in the vulnerable position of having neither knowledge nor control over the mechanisms they employ to know. Algorithms are often *black-boxed* within socio-technical epistemic system; they are either inaccessible or incomprehensible for the users of such systems. These users are therefore either forced to blindly trust these systems and the mechanisms they employ or to withdraw from using systems whose underlying processes they cannot comprehend. A crucial conclusion that I have drawn from this observation was my request for *transparency in socio-technical epistemic systems.* Only in transparent systems can we be responsible, epistemically vigilant knowers.

However trust cannot only be placed and authority not only be ascribed to the mechanisms of closure themselves, but also to different epistemic agents. Hence, a crucial topic in this thesis concerns the *differential assessment of human epistemic agents*. We have seen that despite all contrary rhetoric, the Web is by no means an epistemic democracy where everybody's vote has the same weight. There are examples of unweighted mechanisms, such as the ratings on Amazon.com, where no difference is being made between different users. However, these examples are rare. In the majority of socio-technical epistemic systems, epistemic agents are weighted. Therefore, a crucial socio-epistemological task consists in analyzing these processes of weighting and their consequences.

I have argued that the use of reputational cues to assess the epistemic trustworthiness of epistemic agents is not only common practice, but also epistemically useful. If we do not want to trust blindly, we have to assess epistemic agents differentially for their competency or honesty – or rather for signs of *dis*honesty and *in*competence. In cases where we cannot assess others competence or honesty directly, we may have to rely on proxies of what Kitcher has labelled *unearned authority*, i.e. the "authority that stems from the scientist's social position (either within the community of scientists or in the wider society)" (Kitcher 1993: 315). These indicators related to someone's social positions may sometimes indeed be epistemically useful. Yet as proxies they are only heuristics and as such they are fallible.

Especially feminist scholars have stressed the epistemological and ethical problems that can arise if inadequate proxies for the assessment of trustworthiness are used and epistemic authority is distributed accordingly (e.g. Alcoff 2001, Code 1987, Fricker 2007, Scheman 2001). Hence, although there are epistemic benefits in making differences between epistemic agents, in attributing different amounts of epistemic authority to different agents based on their perceived epistemic trustworthiness, social epistemology has to assess which proxies are valid indicators of epistemic trustworthiness and which ones are not; which reputational cues should be used when ascribing epistemic authority and which ones should not be used. Social epistemology has to assess the relevant criteria and the extent to which the social identity of epistemic agents is relevant for the assessment of their capacities as a knower. It is with respect to these questions, that the entanglement of ethical and epistemological aspects within social epistemology comes to the fore most clearly.

Finally, I have shown that within socio-technical epistemic systems, different closure mechanisms can be combined. Different mechanisms can be employed to end different processes at different stages. *Peer review* has served as one important example to show this nesting of different types of epistemic sociality: peer review can have a selective function (ES^S), while employing integrative (ES^I) and aggregational mechanisms (ES^A) for the formation of a collective judgment. Hence, a crucial field of analysis for future research concerns the ways in which different socio-epistemic mechanisms can best be combined within different socio-technical epistemic systems.

12.4 Limits of this Thesis

The benefits of this thesis are to a certain extent also its limits: its generality and its interdisciplinarity. In this thesis I am offering a new framework for the analysis of epistemic social software – or of socio-technical epistemic systems more generally. It is a framework fundamentally based on works from social epistemology, but combines them with insights from other fields, such as STS and computer ethics. The framework was specifically designed for the analyses of epistemic social software conceived as socio-technical epistemic systems in which numerous human and non-human epistemic agents interact in different forms to reach epistemic goals and produce epistemic products through different mechanisms of closure. Given the socio-epistemological basis of this model, many insights were based on socio-epistemological analyses of socio-epistemic practices in the sciences. Hence, frequent reference and comparisons were made between socio-epistemic practices on the Web and in science. This proximity is one of the reasons why I consider the model I propose suitable for analyses of socio-technical epistemic systems other than epistemic social software. The points made, the questions asked and the issues raised throughout this thesis are as applicable to science as they are to epistemic social software, although the answers may differ. As such, the model proposed can serve as a general socio-epistemological framework that can be fruitfully applied to all socio-epistemological processes that involve multiple agents, although it was developed and is particularly well suited for the analysis of socio-technical epistemic systems, such as epistemic social software.

A shortcoming of such a generic proposal is that there are many blanks. Indeed, I have raised more issues than I have answered. I consider this to be a sign of the fruitfulness of my proposal. However, others may find that the analyses I offer might not go far enough. I agree and hope to be able to continue work on the topics raised in the years to come.

A second related aspect concerns the interdisciplinary nature of this thesis. This comes as well with pros and cons. One of the major shortcomings of such an interdisciplinary work concerns its omissions. Clearly more could be said about social epistemology, about STS, about information and communication technology. This thesis lies at an intersection. Hence, its value lies in making connections, in relating research from different fields to open up new research questions. Its flaw may lie in not going far enough into each direction. Here again I hope to follow some of the routes into the different disciplines touched in more detail in the next years.

Proposing a new frame of reference differs from making an argument within a frame of reference. Hence, despite all my emphasis on processes of closure, this thesis is rather an opening than a closure. With the model I have proposed, I aim at opening a field of research which yet has to be explored: the socio-epistemological analysis of socio-technical epistemic systems. Proposing a comprehensive framework for a new field of research differs from delivering a rigid argument. And if there is one thing I have learned about philosophical method it is that rigor of argumentation is one of the key criteria of quality. If there are any singular arguments which I deliver in this thesis they are the following:

1. Given their prevalence, ubiquity and relevance for contemporary epistemic practices, epistemic social software is an important field to analyze from a socio-epistemological perspective.
2. Social epistemology, while delivering valuable and important insights on numerous aspects of epistemic social software has fallen short of delivering a full-fledged, comprehensive framework for such analyses.
3. Such a framework has to account for insights obtained in different academic fields, most notably from STS, and has to combine an epistemological with an ethical perspective.

These arguments are the red threads that run through this thesis and the model I propose is intended to be such a new framework. I have build this framework upon literature research in different fields and started to validate it with different case studies proposed in the last three chapters. Whether my model will continue to be heuristically fruitful is to be shown in the years to come.

12.5 Future Research

The goal of this thesis was to develop a socio-epistemological framework to assess epistemic social software. So far, there have been only few previous attempts to analyze social software from a socio-epistemological perspective and no comprehensive framework for such an analysis has existed so far. Further, previous analyses have been hampered by a neglect of the crucial role of technology for epistemic practices.

Epistemic practices nowadays *are socio-technical epistemic practices*. Therefore, approaches which neglect the role of technology are inadequate to describe our contemporary epistemic environment.

Accordingly, with this thesis I am opening up a new field of research: the socio-epistemological analysis of epistemic social software. Throughout this thesis, but in particular in Part 3, I have sketched directions for future research and I shall return to these issues to conclude this thesis.

With respect to epistemic social software, the most straightforward request concerns the analyses of more and different types of epistemic social software. While Wikipedia clearly is an important case study from a socio-epistemological perspective, other types of software have to be scrutinized. We have to assess their functionalities, their strengths and weakness, the prerequisites upon which their proper functioning depends, the possible dangers if such prerequisites are not met. We also have to ask different questions. In particular questions concerning the creative and generative function of epistemic social software. The restrictive focus on ICT as a means of knowledge transmission has narrowed the realm of questions that were asked. Hence, we need to broaden our frames again and ask in which ways which types of epistemic content, which types of knowledge, are being created on the Web.

An important research topic concerns the interplay between technical constraints and social norms in socio-technical epistemic systems, the emergence of *socio-technical constraints*. We also have to apprehend the performative nature of this relationship, the question how such social-technical constraints are enacted, fostered or changed over time.

Another topic of crucial relevance concerns the notion of *trust,* and the *interplay between epistemic trust and epistemic vigilance*. Here, empirical questions of how people trust on the Web, whom they trust and for which reasons, have to be combined with normative concerns about the validity of such practices. We have to analyze characteristics of *procedural trust*, how it differs from trust in persons and institution, its pros and cons. Issues around *algorithmic authority* have to be assessed, how it functions and where it malfunctions. Trust in algorithms renders the problem of black-boxing, of accountability and responsibility central. I have argued for *transparency* in socio-technical epistemic systems. Yet, how transparency can be increased without overburdening the users is a question that yet has to be explored. Different forms of *visualizations* have to be conceived, tested and compared to develop new systems that enable their users to be more responsible and accountable knowers.

When it comes to human epistemic agents, we have to assess the processes by which we place trust in them. Whom do we trust for which reasons? Which proxies do we use to assess others' epistemic trustworthiness? Are these processes valid? What are the consequences of using different proxies? What are the pros and cons of ascribing different amounts of trustworthiness to different agents? With respects to these questions, the relationship between ethics and epistemology has to be fathomed.

Epistemic social software is a rich field of research that yet needs to be explored and the questions outlined above merely point into different directions of future research. However, the model I propose is not restricted to epistemic social software. Rather, it is generic enough to deliver a framework for the analysis of socio-technical epistemic systems more broadly conceived. One clear candidate for probing my model is science, because the same questions which I have asked above can also be used to assess epistemic practices within science. This also implies that my model can be compared to other socio-epistemological approaches with respect to its fertility in opening up research questions for the analysis of science. The utility of my framework to analyze issues related to scholarly communication is most straightforward. Hence, I have made reference to projects such as LiquidPub[193], which assess the impact of Web2.0-technologies for scholarly communication and I used topics such as peer review to elucidate the relevance of my model. However, issues of trustworthiness, (procedural) trust, diversity, transparency in socio-technical epistemic systems are not only important to understand scientific publishing and communication, but also to apprehend the knowledge-creative processes within science, the interplay between human and non-human agents in the processes of knowledge creation. Indeed, one may ask whether and to what extent the differentiation into knowledge creation, evaluation and dissemination has to be reconceived to account for epistemic practices in science and beyond.

Finally, it has to be noted that epistemic social software and science are just two very prominent examples of socio-technical epistemic systems. The educational system or industry can also be understood in terms of the interplay between the social, the

[193] http://project.liquidpub.org/ [date of access: 17.04.2010]

epistemic and the technical. Hence, several questions raised in this thesis can certainly also be applied to other societal domains. The field is wide open - the work is yet to be done.

13 References

Adler, J. (1994). "Testimony, Trust, Knowing." The Journal of Philosophy 91 (5): 264-275.

Adler, J. (2006). "Epistemological Problems of Testimony." Stanford Encyclopedia of Philosophy, from http://plato.stanford.edu/entries/testimony-episprob/.

Adomavicius, G. and A. Tuzhilin (2005). "Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions." IEEE Transactions on Knowledge and Data Engineering 17 (6): 734-749.

Agre, P. E. (1997). Toward a Critical Technical Practice: Lessons Learned in Trying to Reform AI. Social Science, Technical Systems, and Cooperative Work: Beyond the Great Divide. G. C. Bowker, S. L. Star, W. Turner and L. Gasser. Mahwah, New Jersey, Erlbaum.

Akrich, M. (1992). The De-scription of Technical Objects. Shaping Technology/Building Society: Studies in Sociotechnical Change. W. E. Bijker and J. Law. Cambridge, MIT Press**:** 205-224.

Akrich, M. and B. Latour (1992). A Summary of a Convenient Vocabulary for the Semiotics of Human and Non-Human Assemblies. Shaping Technology/Building Society: Studies in Sociotechnical Change. W. E. Bijker and J. Law. Cambridge, MIT Press**:** 259-264.

Alcoff, L. M. (2001). On Judging Epistemic Credibility: Is Social Identity Relevant? Engendering Rationalities. N. Tuana and S. Morgen. Albany, SUNY Press**:** 53-80.

Anderson, C. (2006). The Long Tail: Why the Future of Business is Selling Less of More. New York, Hyperion.

Anderson, E. (2009). "Feminist Epistemology and Philosophy of Science." Stanford Encyclopedia of Philosophy, from http://plato.stanford.edu/entries/feminism-epistemology/.

Avesani, P., P. Massa, et al. (2005). "Moleskiing.it: a Trust-Aware Recommender System for Ski Mountaineering." International Journal for Infonomics: 1-10.

Ayres, I. (2007). Super Crunchers: Why Thinking-by-Numbers Is the New Way to Be Smart. New York, Bantam.

Baier, A. C. (1986). "Trust and Antitrust." Ethics 96 (2): 231-260.

Balabanovi, M. and Y. Shoham (1997). "Fab: Content-Based, Collaborative Recommendation." Communications of the ACM 40 (3): 66-72.

Barad, K. (1996). Meeting the Universe Halfway. Realism and Social Constructivism without Contradiction. Feminism, Science, and the Philosophy of Science. L. H. Nelson and J. Nelson. Dordrecht, Holland, Kluwer**:** 161-194.

Barad, K. (1998). Agential Realism: Feminist Interventions in Understanding Scientific Practices. The Science Studies Reader. M. Biaglioli. New York, Routledge**:** 1-11.

Barad, K. (2003). "Posthumanist Performativity: Toward an Understanding of How Matter Comes to Matter." Signs: Journal of Women in Culture and Society 28 (3): 801-831.

Barad, K. (2007). Meeting the Universe Halfway: Quantum Physics and the Entanglement of Matter and Meaning. Durham, Duke University Press.

Barnes, B. and D. Bloor (1982). Relativism, Rationalism, and the Sociology of Knowledge. Rationality and Relativism. R. Ineichen, M. Hollis and S. Lukes. Cambridge, MIT Press.

Bath, C. (2009). De-Gendering Informatischer Artefakte: Grundlagen einer Kritisch-Feministischen Technikgestaltung (PhD Thesis). Fachbereich 3 (Mathematik & Informatik). Bremen, Universität Bremen: 378.

Bayes, T. (1764). "An Essay Towards Solving a Problem in the Doctrine of Chances." Philosophical Transactions of the Royal Society of London 53: 370-418.

Benkler, Y. (2002). "Coase's Penguin, or, Linux and The Nature of the Firm." Yale Law Journal 112 369-446.

Benkler, Y. (2006). The Wealth of Networks: How Social Production Transforms Markets and Freedom. New Haven, Yale University Press.

Benkler, Y. and H. Nissenbaum (2006). "Commons-Based Peer Production and Virtue." Journal of Political Philosophy 14 (4): 394-419.

Biaglioli, M., Ed. (1998). The Science Studies Reader. New York, Routledge.

Bijker, W. E. (1995). Sociohistorical Technology Studies. Handbook of Science and Technology Studies. S. Jasanoff, G. E. Markle, J. C. Peterson and T. Pinch. Thousand Oaks, CA, Sage: 229-256.

Bijker, W. E., T. P. Hughes, et al., Eds. (1987). The Social Construction of Technological Systems. Cambridge, MIT Press.

Bijker, W. E. and J. Law (1992a). General Introduction. Shaping Technology/Building Society: Studies in Sociotechnical Change. W. E. Bijker and J. Law. Cambridge, MIT Press: 1-14.

Bijker, W. E. and J. Law, Eds. (1992b). Shaping Technology/Building Society: Studies in Sociotechnical Change. Cambridge, MIT Press.

Bikhchandani, S., D. Hirshleifer, et al. (1992). "A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades." Journal of Political Economy 100 (5): 992.

Bodmer, W. F. (1985). The Public Understanding of Science (Report). London, The Royal Society.

Bohman, J. (2006). "Deliberative Democracy and the Epistemic Benefits of Diversity." Episteme 3 (3): 175-191.

Borland, J. (2007). "See Who's Editing Wikipedia - Diebold, the CIA, a Campaign." Wired, from http://www.wired.com/politics/onlinerights/news/2007/08/wiki_tracker?currentPage=1.

Bowker, G. C. and S. L. Star (1999). Sorting Things Out: Classification and Its Consequences. Cambridge, MIT Press.

Bragues, G. (2009). "Prediction Markets: The Practical and Normative Possibilities for the Social Production of Knowledge." Episteme 6 (1): 91-106.

Brandom, R. B. (1994). Making it Explicit: Reasoning, Representing, and Discursive Commitment Cambridge, Mass., Harvard University Press.

Butler, J. (1993). Bodies That Matter. New York, Routledge.

Callon, M. (1986). Some Elements of a Sociology of Translation: Domestication of the Scallops and the Fishermen of St. Brieuc Bay. Power, Action, and Belief: A New Sociology of Knowledge? J. Law. London, Routledge and Kegan Paul: 196-233.

Callon, M. (2004). Actor Network Theory. International Encyclopedia of the Social & Behavioral Sciences. N. J. Smelser and P. B. Bates. Amsterdam, Elsevier: 62-66.

Campbell, S. (2001). Memory, Suggestibility, and Social Skepticism. Engendering Rationalities. N. Tuana and S. Morgen. Albany, SUNY Press: 151-173.

Carrier, M. (2004). Knowledge and Control: On the Bearing of Epistemic Values in Applied Science. Science, Values and Objectivity. P. Machamer and G. Wolters. Pittsburgh, University of Pittsburgh Press: 275-293.

Carrier, M. (2008). Science in the Grip of the Economy: On the Epistemic Impact of the Commercialization of Research. The Challenge of the Social and the Pressure of Practice: Science and Values Revisited. M. Carrier, D. Howard and J. A. Kourany. Pittsburgh, University of Pittsburgh Press: 217-234.

Cartwright, N. (1983). How the Laws of Physics Lie. Oxford, Oxford University Press.

Casati, F., M. Marchese, et al. (2009). Is Peer Review Any Good? A Quantitative Analysis of Peer Review (Report). Trento Department of Information Engineering and Computer Science, University of Trento.

Casper, M. (1994). "Reframing and Grounding Nonhuman Agency: What Makes a Fetus an Agent?" American Behavioral Scientist 37 (6): 839-856.

Chi, E. H., B. Suh, et al. (2008). Providing Social Transparency Through Visualizations in Wikipedia. Social Data Analysis Workshop at CHI 2008, Florence, Italy.

Coady, C. A. J. (1992). Testimony. A Philosophical Study. Oxford, Claredon Press.

Coady, D. (2006). "When Experts Disagree." Episteme: A Journal of Social Epistemology 3 (1): 68-79.

Code, L. (1987). Epistemic Responsibility. Hanover, New England, University Press of New England.

Code, L. (1991). What Can She Know? Feminist Theory and Construction of Knowledge. Ithaca, Cornell University Press.

Code, L. (2001). Rational Imaginings, Responsible Knowings: How Far Can You See From Here? . Engendering Rationalities. N. Tuana and S. Morgen. Albany, SUNY Press: 261-282.

Cohen, S. (1987). "Knowledge, Context, and Social Standards." Synthese 73 (1): 3-26.

Collins, H. (1981a). "Stages in the Empirical Programme of Relativism." Social Studies of Science 11: 3-11.

Collins, H. (1981b). "The Experimenters' Regress as Philosophical Sociology." Studies in History and Philosophy of Science A (33): 149-156.

Collins, H. (1985). Changing Order: Replication and Induction in Scientific Practice. London, Sage.

Craig, E. (1990). Knowledge and the State of Nature: An Essay in Conceptual Synthesis. Claredon Press, Oxford.

Crombie, A. C. (1988). "Designed in the Mind: Western Visions of Science, Nature and Humankind." History of Science 24: 1-12.

Crombie, A. C. (1994). Styles of Scientific Thinking in the European Tradition. The History of Argument and Explanation Especially in the Mathematical and Biomedical Sciences and Arts London, Duckworth.

Cronberg, T. (1996). "European TA-discourses - European TA?" Technological Forecasting and Social Change 51 (1): 55-64.

Daston, L. (1993). Historical Epistemology. Questions of Evidence. Proof, Practice, and Persuasion across the Disciplines. J. Chandler, A. I. Davidson and H. D. Harootunian. Chicago, The University of Chicago Press: 282-289.

Daston, L. and P. Galison (2007). Objectivity. New York, Zone Books.

Daukas, N. (2006). "Epistemic Trust and Social Location." Episteme: 109-123.

Davidson, D. (1991/2001). Three Varieties of Knowledge. Subjective, Intersubjective, Objective. D. Davidson. Oxford, Oxford University Press: 205-220.

Dean, C. (2008). If You Have a Problem, Ask Everyone New York Times. New York, The New York Times Company.

Dourish, P., J. Finlay, et al. (2004). Reflective HCI: Towards a Critical Technical Practice. CHI 2004, Vienna, Austria.

Egan, M. E. and J. H. Shera (1949). "Prolegomena to Bibliographic Control." Journal of Cataloging ad Classification 5 (2): 17-19.

Egan, M. E. and J. H. Shera (1952). "Foundations of a Theory of Bibliography." Library Quarterly 22: 125-137.

Elkana, Y. (1986). Anthropologie der Erkenntnis: die Entwicklung des Wissens als episches Theater einer listigen Vernunft. Frankfurt a.M., Suhrkamp.

Fallis, D. (2006). "Social Epistemology and Information Science." Annual Review of Information Science and Technology 40: 475-519.

Fallis, D. (2009). "Introduction: The Epistemology of Mass Collaboration." Episteme 6 (1): 1-7.

Faulkner, P. (1998). "David Hume's Reductionist Epistemology of Testimony." Pacific Philosophical Quaterly 79 (302-313).

Faulkner, P. (2007a). "A Genealogy of Trust." Episteme 4 (3): 305-321.

Faulkner, P. (2007b). "On Telling and Trusting." Mind 116 (464): 875-902.

Feinberg, J. (1985). Sua Culpa. Ethical Issues in the Use of Computers. D. G. Johnson and J. Snapper. Belmont, CA, Wadsworth.

Flanagan, M., D. Howe, et al. (2005). Values in Design: Theory and Practice (Research Report). New York, Hunter College & New York University: 33.

Flanagan, M., D. C. Howe, et al. (2008). Embodying Values in Technology: Theory and Practice. Information Technology and Moral Philosophy. J. v. d. Hoven and J. Weckert. Cambridge Cambridge University Press: 322-353.

Fleck, L. (1980). Entstehung und Entwicklung einer wissenschaftlichen Tatsache: Einführung in die Lehre vom Denkstil und Denkkollektiv. Frankfurt a.M., Suhrkamp.

Floridi, L. (2009). "Web 2.0 vs. the Semantic Web: A Philosophical Assessment." Episteme 6 (1): 25-37.

Foray, D. (2004). The Economics of Knowledge. Cambridge, MIT Press.

Foucault, M. (1970). The Order of Things: An Archeology of the Human Sciences. London, Tavistock.

Foucault, M. (1971). The Archeology of Knowledge. New York, Pantheon.

Foucault, M. (1980). Power/Knowledge: Selected Interviews and Other Writings. New York, Pantheon.

Franck, G. (1996). "Aufmerksamkeit - Die neue Währung." Telepolis, from http://www.heise.de/tp/r4/artikel/2/2003/1.html.

Fricker, E. (1995a). "Critical Notice: Telling and Trusting: Reductionism and Anti-Reductionism in the Epistemology of Testimony." Mind 104 (414): 393-411.

Fricker, E. (2006a). "Second-Hand Knowledge." Philosophy and Phenomenological Research 73 (3): 592-618.

Fricker, E. and D. E. Cooper (1987). "The Epistemology of Testimony." Proceedings of the Aristotelian Society 61 (Supplementary Volumes): 57-106.

Fricker, M. (1998). "Rational Authority and Social Power: Towards a Truly Social Epistemology." Proceedings of the Aristotelian Society 19 (2): 159-177.

Fricker, M. (2006b). "Powerlessness and Social Interpretation." Episteme 3 (1-2): 96-108.

Fricker, M. (2007). Epistemic Injustice. Power and the Ethics of Knowing. Oxford, Oxford University Press.

Friedman, B., Ed. (1997a). Human Values and the Design of Computer Technology. Cambridge, Cambridge University Press.

Friedman, B. (1997b). Introduction. Human Values and the Design of Computer Technology. B. Friedman. Cambridge, Cambridge University Press: 1-18.

Friedman, B., P. H. Kahn, et al. (2006). Value Sensitive Design and Information Systems. Human-Computer Interaction in Management Information Systems: Foundations. P. Zhang and D. Galletta. New York, M.E. Sharpe: 348-372.

Friedman, B. and H. Nissenbaum (1997). Bias in Computer Systems. Human Values and the Design of Computer Technology. B. Friedman. Cambridge, Cambridge University Press: 21-40.

Fuller, S. (1987). "On regulating what is known: A way to social epistemology." Synthese 73 (1): 145-183.

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

Fuller, S. (1992). Social Epistemology and the Research Agenda of Science Studies. Science as Practice and Culture. A. Pickering. Chicago University of Chicago Press: 390-428.

Fuller, S. (1994). "Essay Review: Mortgaging the Farm to Save the (Sacred) Cow." Studies in the History and Philosophy of Science 25 (2): 251-261.

Fuller, S. (2004). "Descriptive vs Revisionary Social Epistemology: The Former as Seen by the Latter." Episteme 1 (1): 23-34.

Fuller, S. (2006). The Philosophy of Science and Technology Studies. New York, Routledge.

Gasser, U. (2006). "Regulating Search Engines: Taking Stock and Looking Ahead." Yale Journal of Law & Technology 9: 124-157.

Gigerenzer, G. (1991). How to Make Cognitive Illusions Disappear: Beyond "Heuristics and Biases". European Review of Social Psychology. W. Stroebe and M. Hewstone. New York, Wiley.

Gilbert, M. (1987). "Modelling Collective Belief." Synthese 73 (1): 185-204.

Gilbert, M. (1989). On Social Facts. Princeton, Princeton University Press.

Gilbert, M. (1994). Remarks on Collective Belief. Social Epistemology. F. F. Schmitt. Lanham, Maryland., Rownman & Littlefield Publishers: 235-256.

Gilbert, M. (2004). "Collective Epistemology." Episteme 1 (2): 95-107.

Giles, J. (2005). "Internet Encyclopaedias Go Head to Head " Nature 438: 900-901.

Golbeck, J. (2006). Generating Predictive Movie Recommendations from Trust in Social Networks. Fourth International Conference on Trust Management, Pisa, Italy.

Golbeck, J. and J. Hendler (2006). FilmTrust: Movie Recommendations from Semantic Web-based Social Networks. IEEE Consumer Communications and Networking Conference, Las Vegas.

Goldhaber, M. H. (1996a, 10.11.2009). "Die Aufmerksamkeitsökonomie und das Netz - Teil I." Telepolis, from http://www.heise.de/tp/r4/artikel/6/6195/1.html.

Goldhaber, M. H. (1996b, 10.11.2009). "Die Aufmerksamkeitsökonomie und das Netz - Teil II." Telepolis, from http://www.heise.de/tp/r4/artikel/6/6200/1.html.

Goldman, A. I. (1987). "Foundations of Social Epistemics." Synthese 73 (1): 109-144.

Goldman, A. I. (1992). Liaisons: Philosophy Meets the Cognitive and Social Sciences. Cambridge, MIT Press.

Goldman, A. I. (2001). "Experts: Which Ones Should You Trust?" Philosophy and Phenomenological Research 63 (1): 85-110.

Goldman, A. I. (2003). Knowledge in a Social World. Oxford, Claredon Press.

Goldman, A. I. (2006). "Social Epistemology." Stanford Encyclopedia of Philosophy Retrieved 29.01.2007, from http://plato.stanford.edu/entries/epistemology-social/.

Goldman, A. I. (2008). The Social Epistemology of Blogging. Information Technology and Moral Philosophy. J. v. d. Hoven and J. Weckert. New York, Cambridge University Press: 11-122.

Goldman, A. I. (2009a). Social Epistemology. Blackwell Companion to Epistemology. J. Dancy, E. Sosa and M. Steup. Oxford, Blackwell: 82-92.

Goldman, A. I. (2009b). Why Social Epistemology is Real Epistemology. D. Pritchard. Oxford, Oxford University Press.

Goldman, A. I. (2009c). Systems-Oriented Social Epistemology. Oxford Studies in Epistemology. T. Gendler and J. Hawthorne.

Grasswick, H. (2006). "Feminist Social Epistemology." Stanford Encyclopedia of Philosophy, from http://plato.stanford.edu/entries/feminist-social-epistemology/.

Hacking, I. (1975). The Emergence of Probability: A Philosophical Study of Early Ideas about Probability, Induction and Statistical Inference. Cambridge, Cambridge University Press.

Hacking, I. (1990). The Taming of Chance. Cambridge, Cambridge University Press.

Hacking, I. (1992a). Statistical Language, Statistical Truth and Statistical Reason: The Self-Authentification of a Style of Scientific Reasoning. Social Dimensions of Science. E. McMullin. Notre Dame, Indiana: 130-157.

Hacking, I. (1992b). "'Style' for Historians and Philosophers." Studies in History and Philosophy of Science 23 (1-20).

Hacking, I. (1992c). The Self-Vindication of the Laboratory Sciences. Science as Practice and Culture. A. Pickering. Chicago University of Chicago Press: 29-64.

Hacking, I. (1995). Immagini Radicalmente Costruzionaliste del Progresso Matematico. Realismo/Antirealismo. A. Pagnini. Florenz, La Nuova Italia: 59-92.

Hacking, I. (2002). "Inaugural Lecture: Chair of Philosophy and History of Scientific Concepts at the Collège de France." Economy and Society 31 (1): 1-14.

Hacking, I. (2004). Historical Ontology. Cambridge, Harvard University Press.

Hacking, I. (2005). Ein Stilbegriff für Historiker und Philosophen. Bilder der Natur - Sprachen der Technik. Nach Feierabend. Zürcher Jahrbuch für Wissensgeschichte. D. Gugerli, M. Hagner, M. Hampeet al. Zürich. 1: 139- 167.

Hansen, A. and C. Clausen (2003). "Social Shaping Perspectives in Danish Technology Assessment." Technology in Society 25 (3): 431-451.

Haraway, D., Ed. (1991a). Simians, Cyborgs and Women: The Reinvention of Nature. New York, Routledge.

Haraway, D. (1991b). A Cyborg Manifesto: Science, Technology, and Socialist-Feminism in the Late Twentieth Century. Simians, Cyborgs

and Women: The Reinvention of Nature. D. Haraway. New York, Routledge: 149-181.

Haraway, D. (1994). "A Game of Cat's Cradle: Science Studies, Feminist Theory, Cultural Studies." Configurations 2 (1): 59-71.

Haraway, D. (1996). Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective. Feminism and Science. E. F. Keller and H. E. Longino. New York, OUP.

Haraway, D. (1997). Modest Witness@Second Millenium. FemaleMan Meets OncoMouse: Feminism and Technoscience. New York, Routledge.

Harding, S. (1991). Whose Science? Whose Knowledge? Thinking from Women's Lives. Milton Keynes, Open University Press.

Harding, S. (2003). Starke Objektivität. Wissen zwischen Entdeckung und Konstruktion. M. Vogel and L. Wingert. Frankfurt a.M., Suhrkamp: 162-192.

Hardwig, J. (1985). "Epistemic Dependence." The Journal of Philosophy 82 (7): 335-349.

Hardwig, J. (1991). "The Role of Trust in Knowledge." The Journal of Philosophy 88 (12): 693-708.

Hidderley, R. and P. Rafferty (2006). "Flickr and Democratic Indexing: Disciplining Desire Lines." Advances in Knowledge Organization 10: 405-411.

Hu, N., P. A. Pavlou, et al. (2009). "Overcoming the J-shaped Distribution of Product Reviews." Communications of the ACM 52 (10): 144-147.

Hughes, T. P. (1983). Networks of Power: Electrification in Western Society, 1880-1930. Baltimore, Johns Hopkins University Press.

Hume, D. (1957). An Enquiry Concerning Human Understanding. Oxford, Clarendon Press.

Ilyes, P. (2006). Zum Stand der Forschung des Internationalen "Science and Technology" (STS)-Diskurses. Frankfurt, Institut für Kulturanthropologie und Europäische Ethnologie, Johann Wolfgang Goethe-Universität: 85.

Introna, L. and H. Nissenbaum (2000). The Public Good Vision of the Internet and the Politics of Search Engines. Preferred Placement - Knowledge Politics on the Web. R. Rogers. Maastricht, Jan van Eyck Akademy: 25-47.

Issing, L. J. and P. Klimsa (1997). Information und Lernen mit Multimedia. Weinheim, Psychologische Verlagsunion.

Joerges, B. (1999). "Do Politics Have Artefacts?" Social Studies of Science 29 (3): 411-431.

Kaplan, A. M. and M. Haenlein (2010). "Users of the World, Unite! The Challenges and Opportunities of Social Media." Business Horizons 53 (1): 59-68.

Keen, A. (2008). The Cult of the Amateur. New York, Doubleday.

Keller, E. F., Ed. (1985). Reflections on Gender and Science. New Haven, Yale University Press.

Keller, E. F. and H. E. Longino, Eds. (1996). Feminism and Science. Oxford, Oxford University Press.

Kelly, P. (2006). "Liberalism and Epistemic Diversity: Mill's Sceptical Legacy." Episteme 3 (3): 248-265.

Kerres, M., D. Euler, et al. (2005). Lehrkompetenz für eLearning-Innovationen in der Hochschule Ergebnisse einer Explorativen Studie zu Massnahmen der Entwicklung von eLehrkompetenz. S CIL-Arbeitsbericht 6. D. Euler and S. Seufert. St. Gallen/Switzerland, Swiss Centre for Innovations in Learning

Kitcher, P. (1990). "The Division of Cognitive Labor." The Journal of Philosophy 87 (1): 5-22.

Kitcher, P. (1993). The Advancement of Science. New York, Oxford University Press.

Kitcher, P. (1994). Contrasting Conceptions of Social Epistemology. Social Epistemology. F. F. Schmitt. Lanham, Maryland., Rownman & Littlefield Publishers: 111-131.

Kling, R. (1992). "Audiences, Narratives and Human Values in Social Studies of Technology." Science, Technology & Human Values 17 (3): 349 - 365.

Knorr-Cetina, K. (1981). The Manufacture of Knowledge. An Essay on the Constructivist and Contextual Nature of Science. Oxford Pergamon Press.

Knorr-Cetina, K. (1984). Die Fabrikation von Erkenntnis - Zur Anthropologie der Naturwissenschaft. Frankfurt, Suhrkamp.

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

Koepsell, D. R. (2003). The Ontology of Cyberspace: Philosophy, Law, and the Future of Intellectual Property. Peru, Illinois, Carus Publishing.

Kornblith, H. (1987). "Some Social Features of Cognition." Synthese 73 (1): 27-41.

Kornblith, H. (1994). A Conservative Approach to Social Epistemology. Social Epistemology. F. F. Schmitt. Lanham, Maryland., Rownman & Littlefield Publishers: 93-110.

Kourany, J. A. (2003). "A Philosophy of Science for the Twenty-First Century." Philosophy of Science 70: 1-14.

Kuhn, T. S. (1962/1970). The Structure of Scientific Revolutions Chicago, University of Chicago Press.

Kusch, M. (2002). Knowledge by Agreement: The Programme of Communitarian Epistemology. Oxford, Oxford University Press.

Kusch, M. (2007). Epistemic Replacement Relativism Defended. EPSA07: Launch of the European Philosophy of Science Association, Complutense University, Madrid, Springer.

Kusch, M. (2009). Social Epistemology. The Routledge Companion to Epistemology. S. Bernecker and D. Pritchard. London, Routledge.

Kusch, M. (2009b). Kripke's Wittgenstein, *On Certainty*, and Epistemic Relativism. The Later Wittgenstein on Language. D. Whiting. Houndmills, Basingstoke, Palgrave Macmillan**:** 213-229.

Kusch, M. (2009c). Testimony and the Value of Knowledge. Epistemic Value. A. Haddock, A. Millar and D. Pritchard. Oxford, Oxford University Press**:** 60-94.

Kusch, M. (to appear). Social Epistemology. The Routledge Companion to Epistemology. S. Bernecker and D. Pritchard. London, Routledge.

Lackey, J. (2007). "Introduction: Perspectives on Testimony." Episteme 4 (3): 233-237.

Lackey, J. and E. Sosa, Eds. (2006). The Epistemology of Testimony. Oxford, Claredon Press.

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

Langlois, G. M. (2005). "Networks and Layers: Technocultural Encodings of the World Wide Web." Canadian Journal of Communication 30 (4).

Latour, B. (1987). Science in Action. Cambridge, Harvard University Press.

Latour, B. (1992). Where Are the Missing Masses? The Sociology of a Few Mundane Artifacts. Shaping Technology/Building Society: Studies in Sociotechnical Change. W. E. Bijker and J. Law. Cambridge, MIT Press**:** 225-258.

Latour, B. (1996). Der Berliner Schlüssel. Berlin, Akademie Verlag.

Latour, B. (1999). On Recalling ANT. Actor Network Theory and After J. Law and J. Hassard. Oxford, UK Blackwell**:** 15-25.

Latour, B. (2000). Die Hoffnung der Pandora. Frankfurt a.M., Suhrkamp Verlag.

Latour, B. (2005). Reassembling the Social. An Introduction to Actor-

Network-Theory. Oxford, Oxford University Press.

Latour, B. and S. Woolgar (1986). Laboratory Life: The Construction of Scientific Facts. Princeton, Princeton University Press.

Law, J. (1992/2001). "Notes on the Theory of the Actor Network: Ordering, Strategy, and Heterogeneity." from http://www.lancs.ac.uk/fass/sociology/papers/law-notes-on-ant.pdf.

Law, J. (2004). After Method: A Mess in Social Science Research. New York, Routledge.

Law, J. and J. Hassard, Eds. (1999). Actor Network and After. Oxford, Blackwell.

Lehrer, K. (1987). "Personal and Social Knowledge." Synthese 73 (87-107).

Lehrer, K. (1990a). Metamind. Oxford, Claredon Press.

Lehrer, K. (1990b). Theory of Knowledge. Boulder, Westview.

Lehrer, K. and C. Wagner (1981). Rational Consensus in Science and Society. Dordrecht, Reidel.

Leplin, J. (1994). "Critical Note: Philip Kitcher's *The Advancement of Science: Science Without Legend, Objectivity Without Illusion*." Philosophy of Science 61: 666-671.

Lessig, L. (2001). The Future of Ideas: the Fate of the Commons in a Connected World. New York, Random House.

Lessig, L. (2006). Code: And Other Laws of Cyberspace, Version 2.0. New York, Basic Books.

Levien, R. and A. Aiken (1998). Attack-Resistant Trust Metrics for Public Key Certification. 7th Conference on USENIX Security Symposium, San Antonio, Texas

Linestone, H. A. and M. Turoff (2002). The Delphi Method: Techniques and Applications, Addison-Wesley Publishing Company.

Lipton, P. (1998). "The Epistemology of Testimony." Studies in the History and Philosophy of Science 29 (A): 1-31.

Locke, J. (1961). An Essay Concerning Human Understanding. London, Dent.

Loftus, E. F. and H. G. Hoffman (1989). "Misinformation and Memory: The Creation of Memory." Journal of Experimental Psychology: General 118: 100-104.

Loftus, E. F. and J. C. Palmer (1974). "Reconstruction of Automobile Destruction." Journal of Verbal Learning and Verbal Behaviour 13: 585-589.

Longino, H. E. (1990). Science as Social Knowledge: Values and Objectivity in Scientific Inquiry. Princeton, Princeton University Press.

Longino, H. E. (1994). The Fate of Knowledge in Social Theories of Science. Social Epistemology. F. F. Schmitt. Lanham, Maryland., Rownman & Littlefield Publishers**:** 135-157.

Longino, H. E. (1995). "Gender, Politics, and the Theoretical Virtues." Synthese 104 (383-397).

Longino, H. E. (1996). Subjects, Power, and Knowledge: Description and Prescription in Feminist Philosophies of Science. Feminism and Science. E. F. Keller and H. E. Longino. Oxford, Oxford University Press**:** 264-279.

Longino, H. E. (2002c). The Fate of Knowledge. Princeton, Princeton University Press.

Longino, H. E. (2005). "Circles of Reason: Some Feminist Reflections on Reason and Rationality." Episteme 2 (1): 79-88.

Longino, H. E. and K. Lennon (1997). "Feminist Epistemology as a Local Epistemology." Proceedings of the Aristotelian Society Supplementary Volumes 71: 19-35+37-54.

MacKenzie, D. and J. Wajcman, Eds. (1999). The Social Shaping of Technology. Buckingham, Philadelphia, Open University Press.

Mager, A. (2009). "Mediated Health. Sociotechnical Practices of Providing and Using Online Health Information." New Media & Society 11 (7): 1123-1142

Magnus, P. D. (2009). "On Trusting Wikipedia." Episteme 6 (1): 74-90.

Malone, T. W. (1997). Commentary on Suchman Article and Winograd Response. Human Values and the Design of Computer Technology. B. Friedman. Cambridge, Cambridge University Press: 115-116.

Mannheim, K. (1936). Ideology and Utopia. New York, Harcourt, Brace and World.

Massa, P. and P. Avesani (2006). Trust-Aware Bootstrapping of Recommender Systems. ECAI, Riva del Garda, Italy.

Massa, P. and P. Avesani (2007). "Trust Metrics on Controversial Users: Balancing between Tyranny of the Majority and Echo Chambers." International Journal on Semantic Web and Information Systems (IJSWIS) 3 (1): 39-64.

Massa, P. and B. Bhattacharjee (2004). Using Trust in Recommender Systems: an Experimental Analysis. iTrust2004.

Mathiesen, K. (2005). "The Epistemic Features of Group Belief." Episteme 2 (3): 161-175.

McLeod, C. (2006). "Trust." Stanford Encyclopedia of Philosophy Retrieved 20.10.2009, from http://plato.stanford.edu/entries/trust/.

Merriam-Webster-Online-Dictionary (2009). Value. Merriam-Webster Online Dictionary.

Merton, R. K. (1973). The Sociology of Science: Theoretical and Empirical Investigations. Chicago, University of Chicago Press.

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

Mill, J. S. (1859). On Liberty. London, John W. Parker & Son.

Mitcham, C. (2005). Values and Valuing. Encyclopedia of Science, Technology, and Ethics. C. Mitcham. Detroit, MI, Macmillan Reference. 4.

Nelson, L. H. (1993). Epistemological Communities. Feminist Epistemologies. L. Alcoff and E. Potter. New York, Routledge: 121-159.

Nielsen, J. (1999). Designing Web Usability. Thousand Oaks, Peachpit Press.

Nissenbaum, H. (1997). Accountability in a Computerized Society. Human Values and the Design of Computer Technology. B. Friedman. Cambridge, Cambridge University Press: 41-64.

Nissenbaum, H. (2005). Values in Technical Design. Encyclopedia of Science, Technology and Ethics. C. Mitcham. New York, Macmillan: lxvi-lxx.

Nonaka, I. and R. Toyama (2003). "The Knowledge-Creating Theory Revisited: Knowledge Creation as a Synthesizing Process." Knowledge Management Research and Practice 1: 2-10.

Nucci, L. (1996). Morality and the Personal Sphere of Actions. Values and Knowledge E. S. Reed, E. Turiel and T. Brown. Mahwah, New Jersey, Lawrence Erlbaum: 41-60.

O'Brien, J. (2006). "The Race to Create a 'Smart' Google." Fortune, from http://money.cnn.com/magazines/fortune/fortune_archive/2006/11/27/8394347/index.htm?section=money_latest.

Ober, J. (2006). "From Epistemic Diversity to Common Knowledge: Rational Rituals and Cooperation in Democratic Athens." Episteme 3 (3): 214-233.

Oreskes, N. (2008). "The Devil is in the (Historical) Details: Continental Drift as a Case of Normatively Appropriate Consensus?" Perspectives on Science 16 (3): 253-264.

Origgi, G. (to appear). Designing Wisdom Through the Web. The Passion of Ranking. Collective Wisdom. J. Elster and H. Landermore. Cambridge, Cambridge University Press.

Origgi, G. and J. Simon (2010). On the Epistemic Value of Reputation: The Place of Ratings and Reputational Tools in Knowledge Organization. Eleventh International ISKO Conference 2010: Paradigms and conceptual systems in KO, Rome.

Oudshoorn, N. and T. Pinch, Eds. (2005a). How Users Matter: The Co-Construction of Users and Technology Inside Technology. Cambridge/ Mass., MIT Press.

Oudshoorn, N. and T. Pinch (2005b). How Users and Non-users Matter. How Users Matter: The Co-Construction of Users and Technology N. Oudshoorn and T. Pinch. Cambridge, MIT Press: 1-28.

Page, S. E. (2001). "Problem Solving by Heterogeneous Agents." Journal of Economic Theory 97: 123-163.

Page, S. E. (2007). The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies. Princeton Princeton University Press.

Palfrey, J. and U. Gasser (2008). Born Digital: Understanding the First Generation of Digital Natives. New York, Basic Books.

Pinch, T. J. and W. E. Bijker (1984). "The Social Construction of Facts and Artifacts: Or How the Sociology of Science and the Sociology of Technology Might Benefit Each Other." Social Studies of Science: 399-441.

Pinch, T. J. and W. E. Bijker (1987). The Social Construction of Facts and Artifacts: Or How the Sociology of Science and the Sociology of Technology Might Benefit Each Other. The Social Construction of Technological Systems. W. E.Bijker, T. P. Hughes and T. J. Pinch. Cambridge MA, London, England, MIT Press: 17-50.

Polanyi, M. (1985). Implizites Wissen. Frankfurt a.M., Suhrkamp Verlag.

Porter, T. M. (1995). Trust in Numbers: The Pursuit of Objectivity in Science and Public Life. Princeton, Princeton University Press.

Pritchard, D. (2004). The Epistemology of Testimony. Epistemology. E. Sosa and E. Villanueva. Boston, Blackwell Publishing: 326-348.

Quinton, A. (2004). "Two Kinds of Social Epistemology." Episteme 1 (1): 7-9.

Rainie, L. (2007). 28% of Online Americans Have Used the Internet to Tag Content. Forget Dewey and His Decimals, Internet Users are Revolutionizing the Way We Classify Information – and Make Sense of It. PEW Internet & American Life Project. Washington, DC. , PEW Research Center.

Rawls, J. (1971). A Theory of Justice. Cambridge, Massachusetts, Belknap Press.

Reichenbach, H. (1938). Experience and Prediction. Chicago, University of Chicago Press.

Reid, T. (1983). Inquiry and Essays. Indiannapolis, Hackett.

Reinhart, M. (2009). "Peer Review of Grant Applications in Biology and Medicine. Reliability, Fairness, and Validity." Scientometrics 81 (3): 789-809.

Resnick, P. and H. R. Varian (1997). "Recommender Systems." Communications of the ACM 40 (3): 56-58.

Rogers, R. (2009). The Googlization Question, and the Inculpable Engine. Deep Search: The Politics of Search Engines. K. Becker and F. Stalder. Edison, NJ, Transaction Publishers.

Röhle, T. (2009). Dissecting the Gatekeepers. Relational Perspectives on the Power of Search Engines. Deep Search: The Politics of Search Engines beyond Google. K. Becker and F. Stalder. Edison, NJ, Transaction: 117-132

Rommes, E. (2002). Gender Scripts and the Internet. The Design and Use of Amsterdam's Digital City. Enschede, Twente University Press.

Rouse, J. (1996). Engaging Science: How to Understand Its Practices Philosophically. Ithaka, Cornell University Press.

Sanger, L. M. (2009). "The Fate of Expertise after Wikipedia." Episteme 6 (1): 52-73.

Scheman, N. (2001). Epistemology Resuscitated: Objectivity as Trustworthiness. Engendering Rationalities. N. Tuana and S. Morgen. Albany, SUNY Press: 23-52.

Schmitt, F. F. (1987). "Introduction." Synthese 73 (1): 1-2.

Schmitt, F. F. (1987b). "Justification, Sociality, and Autonomy." Synthese 73 (1): 43-85.

Schmitt, F. F. (1994a). Socializing Epistemology: An Introduction Through Two Sample Issues. Social Epistemology. F. F. Schmitt. Lanham, Maryland., Rownman & Littlefield Publishers: 1-27.

Schmitt, F. F., Ed. (1994b). Social Epistemology: The Social Dimensions of Knowledge. Studies in Epistemology and Cognitive Theory. Lanham, Maryland., Rownman & Littlefield Publishers.

Schmitt, F. F. (1994c). The Justification of Group Beliefs. Social Epistemology. F. F. Schmitt. Lanham, Maryland., Rownman & Littlefield Publishers: 257-287.

Schmitt, F. F. (1994d). Socializing Epistemology: The Social Dimensions of Knowledge. Lanham, Maryland., Rownman & Littlefield Publishers.

Schön, D. A. (1983). The Reflective Practitioner. New York, Basic Books.

Schot, J. and A. Rip (1997). "The Past and Future of Constructive Technology Assessment." Technological Forecasting and Social Change 54: 251-268.

Schrage, M. (2008, 12.02.2010). "Im Netz der Empfehlungen " Technology Review, from http://www.heise.de/tr/artikel/Im-Netz-der-Empfehlungen-275252.html.

Shannon, C. E. (1948). "A Mathematical Theory of Communication." Bell System Technical Journal 27: 379-423; 623-656.

Shapin, S. (1994). A Social History of Truth: Civility and Science in Seventeenth-Century England. Chicago, University of Chicago Press.

Shapin, S. and S. Schaffer (1985). Leviathan and the Air-Pump: Hobbes, Boyle, and the Experimental Life. Princeton, NJ. , Princeton University Press.

Shera, J. H. (1950). Classification as the Basis of Bibliographic Organization. Bibliographic Organization: 15th Annual Conference of the Graduate Library School Chicago, University of Chicago Press.

Shera, J. H. (1970). Sociological Foundations of Librarianship. New York, Asia Publishing House.

Shirky, C. (2005). "Ontology is Overrated: Categories, Links, and Tags." from http://www.shirky.com/writings/ontology_overrated.html.

Shirky, C. (2009). "A Speculative Post on the Idea of Algorithmic Authority." from http://www.shirky.com/weblog/2009/11/a-speculative-post-on-the-idea-of-algorithmic-authority/.

Simon, J. (2007). Social Epistemology and Communication: Between Calculation, Consensus and Communities. Proceedings of LogKCA-07 - ILCLI International Workshop on Logic and Philosophy of Knowledge, Communication and Action. X. Arrazola and J. M. Larrazabal. Donostia, The University of the Basque Country Press: 369-391.

Simon, J. (2008). Knowledge and Trust in Epistemology and Social Software/ Knowledge Technologies. Culture and identity in knowledge organization: Proceedings of the Tenth International ISKO Conference. C. Arsenault and J. T. Tennis. Montréal, Canada, Würzburg: Ergon: 216-221.

Simon, J. (2009). MyChoice & Traffic Lights of Trustworthiness: Where Epistemology Meets Ethics in Developing Tools for Empowerment and Reflexivity. Proceedings of the 8th International Conference of Computer Ethics, Corfu, Nomiki Bibliothiki.

Simpson, J. A. and E. S. C. Weiner (1989). Value. Oxford English Dictionary. J. A. Simpson and E. S. C. Weiner. Oxford, Clarendon Press.

Smetana, J. G. (1983). "Socio-Cognitive Development: Domain Distinctions and Coordinations." Developmental Review 3: 131-147.

Smith, S. M. and S. E. Blankenship (1989). "Incubation Effects." Bulletin of the Psychonomic Society 27 (4): 311-314.

Solomon, M. (1994). A More Social Epistemology. Social Epistemology. F. F. Schmitt. Lanham, Maryland., Rownman & Littlefield Publishers: 217-233.

Solomon, M. (2001). Social Empiricism. Cambridge, MIT Press.

Solomon, M. (2006). "Norms of Epistemic Diversity." Episteme 3 (1-2): 23-36.

Sperber, D., F. Clément, et al. (to appear). "Epistemic Vigilance." Mind.

Spinuzzi, C. (2003). "More than One, Less than Many: A Review of Three “Post-ANT” Books " Currents in Electronic Literacy 7.

Suchman, L. A. (1987). Plans and Situated Actions: The Problem of Human-Machine Communication. New York, Cambridge University Press.

Suchman, L. A. (1997). Do Categories have Politics? The Language/Action Perspective Reconsidered. Human Values and the Design of Computer Technology. B. Friedman. Cambridge, Cambridge University Press: 91-106.

Suchman, L. A. (2007/2009). Human-Machine Reconfigurations. Plans and Situated Actions. Cambridge, Cambridge University Press.

Suchman, L. A., J. Blomberg, et al. (1999). "Reconstructing Technologies as Social Practice." American Behavioral Scientist (43): 392-408.

Suh, B., E. H. Chi, et al. (2008). Lifting the Veil: Improving Accountability and Social Transparency in Wikipedia with WikiDashboard. 26th Annual ACM Conference on Human Factors in Computing Systems, Florence, Italy, NY: ACM.

Sunstein, C. R. (2002). Republic.com. Princeton, Princeton University Press.

Sunstein, C. R. (2006). "Deliberating Groups versus Prediction Markets (or Hayek's Challenge to Habermas)." Episteme 3 (3): 192-213.

Surowiecki, J. (2004). The Wisdom of Crowds: Why the Many are Smarter than the Few and how Collective Wisdom Shapes Business, Economies, Societies, and Nations. New York, Random House.

Taleb, N. N. (2007). The Black Swan: The Impact of the Highly Improbable. New York, Random House.

Tapscott, D. and A. D. Williams (2006). Wikinomics: How Mass Collaboration Changes Everything. New York, Portfolio.

Tatnall, A. (2003). Actor-Network Theory as a Socio-Technical Approach to Information Systems Research. Socio-Technical and Human Cognition Elements of Information Systems, IGI Publishing: 266-283.

Thagard, P. (1997a). "Collaborative Knowledge." Noûs 31 (2): 242-261.

Thagard, P. (1997b). "Internet Epistemology: Contributions of New Information Technologies to Scientific Research." from http://cogsci.uwaterloo.ca/Articles/Pages/Epistemology.html.

Thagard, P. (1998a). "Ulcers and Bacteria I: Discovery and Acceptance." Studies in History and Philosophy of Science. Part C: Studies in History and Philosophy of Biology and Biomedical Sciences 29: 107-136.

Thagard, P. (1998b). "Ulcers and Bacteria II: Instruments, Experiments, and Social Interactions." Studies in History and Philosophy of Science. Part C: Studies in History and Philosophy of Biology and Biomedical Sciences 29 (2): 317-342

Thagard, P. (1998c). Computation and the Philosophy of Science. The Digital Phoenix: How Computers are Changing Philosophy. T. W. Bynum and J. H. Moor. Oxford, Blackwell: 48-61.

Tollefsen, D. P. (2007). "Group Testimony." Social Epistemology 21 (3): 299-311.

Tollefsen, D. P. (2009). "Wikipedia and the Epistemology of Testimony." Episteme 6 (1): 8-24.

Toulmin, S. (1953). The Philosophy of Science. London, Hutchinson.

Tuana, N., Ed. (1989). Feminism and Science. Race, Gender, and Science. Bloomington, Indiana University Press.

Tuomela, R. (2004). "Group Knowledge Analyzed." Episteme 1 (2): 109-127.

Van House, N. (2003). "Science and Technology Studies and Information Studies." Annual Review of Information Science and Technology (ARIST) 38: 3-86.

Wajcman, J. (2002). "Addressing Technological Change: The Challenge to Social Theory." Current Sociology 50: 347-363.

Wajcman, J. (2004). TechnoFeminism. Cambridge, UK, Polity Press.

Walsham, G. (1997). Actor-Network Theory and IS Research: Current Status and Future Prospects. Proceedings of the IFIP TC8 WG 8.2 International Conference on Information Systems and Qualitative Research. Philadelphia, Pennsylvania, United States, Chapman & Hall, Ltd.: 466-480.

Waters, N. L. (2007). "Why You Can't Cite Wikipedia in My Class." Communications of the ACM 50 (9): 15-17.

Welbourne, M. (1986). The Community of Knowledge. Aberdeen, University Press.

Welbourne, M. (2002). "Is Hume Really a Reductivist?" Studies In History and Philosophy of Science Part A 33 (2): 407-423.

Williams, R. and D. Edge (1996). "The Social Shaping of Technology." Research Policy 25 (6): 865-899.

Winner, L. (1980). "Do Artifacts Have Politics?" Daedalus 109 (1): 121-136.

Winner, L. (1993). "Upon Opening the Black Box and Finding It Empty: Social Constructivism and the Philosophy of Technology " Science, Technology, and Human Values 18: 362-378.

Winograd, T. (1997). Categories , Disciplines, and Social Coordination. Human Values and the Design of Computer Technology. B. Friedman. Cambridge, Cambridge University Press: 107-114.

Wolfers, J. and E. Zitzewitz. (2007). "Interpreting Prediction Market Prices as Probabilities." from http://bpp.wharton.upenn.edu/jwolfers/Papers/InterpretingPredictionMarketPrices.pdf.

Wray, K. B. (2009). "The Epistemic Cultures of Science and Wikipedia: A Comparison." Episteme 6 (1): 38-51.

Wyatt, S. (2005). Non-Users Also Matter: The Construction of Users and Non-Users of the Internet. How Users Matter: The Co-Construction of Users and Technology N. Oudshoorn and T. Pinch. Cambridge/ Mass., MIT Press: 67-80.

Wynne, B. (1995). Public Understanding of Science. Handbook of Science and Technology Studies S. Jasanoff, G. E. Markle, J. C. Peterson and T. Pinch. Thousand Oaks, Sage Publications: 361-391.

Zandonade, T. (2004). "Social Epistemology from Jesse Shera to Steve Fuller." Library Trends 52 (4): 810-832.

Ziegler, C. and G. Lausen (2004). Spreading Activation Models for Trust Propagation. IEEE international Conference on E-Technology, E-Commerce and E-Service Washington, DC.

Ziman, J. (2003). "Non-Instrumental Roles of Science." Science and Engineering Ethics 9: 17-27.

14 Curriculum Vitae Judith Simon

Academic Affiliations	
Since 05/2009	Research fellow at the Institut Jean Nicod, Ecole Normale Supérieure, Paris, France
03/2008-09/2008	Visiting scholar at the Forum on Contemporary Europe, Stanford University, USA.
01/2006-05/2009	University assistant at the Department of Philosophy, University of Vienna, Austria.
05/2005-01/2006	University assistant at the Department of Philosophy of Science, University of Vienna, Austria
08/2003-04/2005	Research assistant of the Research Center Juelich at the Max-Delbrueck Center for Molecular Medicine, Research Group: Bioethics & Science Communication, Berlin, Germany.

Education	
since 2005	PhD-Program in Philosophy at the University of Vienna, Austria
2002	Master in Psychology at the Free University of Berlin, Germany
1999-2002	Free University Berlin (Germany), Studies in Psychology and Media Studies
1998-1999	Wilfrid-Laurier University, Waterloo (Canada), Studies in Psychology and English Language & Literature
1996-2000	Philipps-University Marburg (Germany), Studies in Psychology, English Language & Literature, German Language & Literature
1996	Abitur (Stiftschule St. Johann, Amoeneburg, Germany)

Scholarships	
05-12/2009	Research Scholarship F-605, University of Vienna
03 – 09/2008	Stanford Advanced Graduate Exchange Program; visiting scholar at the Forum on Contemporary Europe, Stanford University, USA.
05-06/2007	CEEPUS Teacher Mobility Grant by the Austrian Exchange Service (ÖAW) for teaching at the Department of Philosophy, University of Ljubljana/Slovenia
09/1998-05/1999	Student-Exchange Scholarship by the University of Marburg to study at Wilfrid Laurier University, Canada

Publications:

Origgi, G. & Simon, J. (2010). On the Epistemic Value of Reputation: The Place of Ratings and Reputational Tools in Knowledge Organization. In: Mazzocchi, F. & Gnoli, C. (Eds.). Proceedings of the Eleventh International ISKO Conference (Rome, Feb. 23-26, 2010), *Advances in knowledge organization, 12*. Würzburg: Ergon.

Simon, J. (2009). MyChoice & Traffic Lights of Trustworthiness: Where Epistemology Meets Ethics in Developing Tools for Empowerment and Reflexivity. In: Bottis, Maria; Grodzinsky, Frances; Tavani, Herman; Vlamos, Panayotis (2009), *Proceedings of the 8th International Conference of Computer Ethics*: Nomiki Bibliothiki.

Simon, J. (2008). Knowledge and Trust in Epistemology and Social Software/ Knowledge Technologies. In: Arsenault, C., & Tennis, J. T. (Eds.). (2008). Culture and identity in knowledge organization: Proceedings of the Tenth International ISKO Conference (Montréal, Canada, August 5-8, 2008). *Advances in knowledge organization, 11.* Würzburg: Ergon, p. 216-221.

Simon, J. (2007). Social Epistemology and Communication: Between Calculation, Consensus and Communities. In: X. Arrazola & J.M. Larrazabal (Eds.) *Proceedings of LogKCA-07 - ILCLI International Workshop on Logic and Philosophy of Knowledge, Communication and Action 2007*, Donostia: The University of the Basque Country Press, p. 369-391.

Simon, J. (2007). Probing Concepts: Knowledge and Information as Boundary Objects in Interdisciplinary Discourse. In: B. Rodríguez Bravo & L. Alvite Díez (Eds.) Interdisciplinarity and Transdisciplinarity in the Organization of Scientific Knowledge. *Proceedings of 8th ISKO Spain Congress*. León: Universidad de León, Secretariado de Publicaciones, pp. 147-155.

Simon, J. (2006). Interdisciplinary Knowledge Creation - Using Wikis in Science. In: G. Budin, C. Swertz, and K. Mitgutsch (Eds.), Knowledge organization for a global learning society - Proceedings of the Ninth International ISKO Conference (Vienna, Austria, July 4-7, 2006). *Advances in knowledge organization, 10.* Würzburg: Ergon, pp. 123-130.

Simon, J. (2006). Review of: C. S. Hermann, M. Pauen, J. Rieger & S. Schicktanz (Hrsg.):"Bewusstsein", Wilhelm Fink: Stuttgart, 2005. *Universitas, 12*, 1303-1308.

Simon, J. (2006). Review of: S. Ehm & S. (Hrsg.):"Körper als Maß? Biomedizinische Eingriffe und ihre Auswirkungen auf Körper- und Identitätsverständnisse", Hirzel Verlag: Stuttgart, 2006. *Universitas, 8*, 864-867

Pompe, S., Simon, J., Wiedemann, P. M. & Tannert, C. (2005) Future trends and challenges in pathogenomics: A foresight study. *EMBO Reports, 6* (7), 600 -605.

Wiedemann, P., Simon, J., Schicktanz, S. & Tannert, C. (2004): The future of stem cell research in Germany. *EMBO Reports, 5,* 927-931.

15 Abstract

In recent years new applications emerged on the Web which received the labels *Web2.0* or *social software.* In many of these applications people are engaged in *epistemic activities*, such as the dissemination, organization or creation of knowledge. The goal of this thesis is to analyze the *epistemological* relevance of such epistemic social software. Because communication and interaction between multiple agents seems to be the key to understand the epistemic processes within such systems, *social epistemology*, the philosophical discipline exploring the ways and the extent to which knowledge is social, was chosen as a theoretical framework. However, none of the existing comprehensive social epistemologies delivers a sufficient framework to analyze epistemic social software. Therefore, I have developed a new socio-epistemological framework to analyze epistemic social software which is rooted in socio-epistemological discourse, but amends it with insights from the field of *Science and Technology Studies* (STS).

My framework is founded on a tripartite classification of socio-technical epistemic system based on the mechanisms they employ to close socio-epistemic processes. These three mechanisms are *integration, aggregation and selection*. With this classification I do not aim at reducing the differences between systems to their mechanisms of closure. However, I argue that the classification based on this indicator is heuristically fruitful. Systems employing different mechanisms of closure depend on different social, technical and epistemic prerequisites, have different strengths and weaknesses and are optimal for different epistemic tasks. My model puts a fact into the focus that has been neglected so far in social epistemology: the technical and its relationship to the social and the epistemic. Since most epistemic practices are nowadays pervaded by technologies, such a consideration of the role of technologies in these practices seems to be indispensable for *any* social epistemology that aims at being not only normatively appropriate, but also empirically adequate.

16 Zusammenfassung

In den letzten Jahren sind eine Reihe neuer Anwendungen im Internet entstanden, die zumeist als *Web2.0 oder social software* bezeichnet werden. Viele dieser Anwendungen sind gekennzeichnet durch die Einbindung mehrerer Agenten in Prozesse zur Verbreitung, Organisation und Herstellung von *Wissen*. Das Ziel der vorliegenden Dissertation besteht in der Analyse der epistemologischen Relevanz dieser *epistemischen social software Anwendungen*. Da die Kommunikation und Interaktion zwischen mehreren Agenten deren Schlüsselmerkmal darstellt, bildet die *Soziale Erkenntnistheorie* als philosophische Disziplin, welche die Weisen untersucht, in denen Wissen sozial bedingt ist, die theoretische Grundlage für die Analyse der epistemischen Prozesse innerhalb dieser Systeme. Weil bisher keine soziale Erkenntnistheorie eine ausreichende Theorie für die Analyse *epistemischer social software* zur Verfügung stellen konnte, habe ich die Grundlagen für ein neues sozio-epistemisches Model entwickelt, welches zwar im sozio-epistemologischen Diskurs verankert ist, jedoch um Erkenntnisse aus dem Feld der *Science and Technology Studies* (STS) erweitert wurde.

Dieses Model gründet in der Klassifikation von sozio-technischen epistemischen Systemen anhand unterschiedlicher Mechanismen der Schließung, welche zur Beendigung sozio-epistemischer Prozesse verwendet werden. Diese Klassifikation anhand der drei Schließungsmechanismen *Integration*, *Aggregation* und *Selektion* zielt nicht auf die Einebnung der Differenzen zwischen sozio-technischen epistemischen Systemen, vielmehr liegt ihr Wert in ihrer heuristischen Fruchtbarkeit, darin Differenzen aufzumachen. Systeme, welche unterschiedliche Schließungsmechanismen nutzen, sind gebunden an unterschiedliche soziale, technische und epistemische Voraussetzungen, sie haben unterschiedliche Stärken und Schwächen und eignen sich daher für unterschiedliche epistemische Aufgaben. Das von mir entwickelte Modell lenkt dabei die Aufmerksamkeit auf ein bislang weitgehend in der sozialen Erkenntnistheorie vernachlässigtes Thema: das Technische und seine Beziehung zum Sozialen und zum Epistemischen. Da die meisten epistemischen Praktiken heute durchdrungen sind von Technologie, ist deren Berücksichtigung von entscheidender Bedeutung für jede soziale Erkenntnistheorie, die beansprucht, nicht nur normativ angemessen, sondern auch empirisch adäquat zu sein.