Computing Genomic Science: Bioinformatics and Standardisation in Proteomics

by

Jamie Lewis

Cardiff University

This thesis is submitted to the University of Wales in fulfilment of the requirements for the degree of

DOCTOR IN PHILOSOPHY

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Acknowledgments

Unfortunately I have a limited word count so I cannot thank everyone I would like to. However, there are several people without whose time, support, generosity and faith I would not have reached this point. First, I am greatly indebted to Sara Delamont and Paul Atkinson without whom I would never have done a Ph.D. They have been a source of great inspiration to me. Special thanks are also reserved for Beck, Pam and Helen G for spending hours checking my references and proof reading, and Andy, Ingrid and Neil for their invaluable academic input. They, together with Alex, Amy, Cathy, Helen B, Jackie, Jasmin, Jo, Kelly, Liz, Maggie, Mark, Martin, Mel, Nina and Olly have been a source of support, friendship, motivation and help.

I would also like to thank my supervisors Peter and Ruth, my examiners Nik and Steve, and the KES group for their critique and comments. Thanks must also be reserved for the ESRC for their appreciated financial contribution and Cesagen members for their kindness and encouragement.

Last but certainly not least, I would like to thank my family for their love and support. I particularly would like to dedicate this thesis to my Mum and Dad who have always encouraged me to pursue my own career and have supported me through a few health problems.

Efallai nad yw hwn o safon Da Vinci, ond fe wnes i fy ngorau glas!

Summary

Science is divided and compartmentalised into distinct areas of research. As science develops new research areas emerge and nurture new technologies, new methodological approaches, new disciplines and new research communities. These demarcations are *socially constructed* spaces that impose a sense of order on science by authenticating the new forms of knowledge that surface. Simply stated, the specific research areas and the social relations contained within them, enable science to progress in a proficient, communal, and sometimes cumulative manner. In this sense the constructed *boundaries* can be viewed as a set of ordering devices.

The mapping of the Human Genome was a significant technical event that reordered biological activity by creating a number of these new socially constructed spaces. This celebrated scientific achievement helped yield a number of emerging 'omic' disciplines, numerous innovative high-throughput technologies, and a myriad of embryonic scientific communities, each with its own distinct identity. In this thesis the Human Genome Project is viewed as the genomic stage of the omic revolution or stage one. The period directly after the sequencing has been coined the post-genomic era and this is described in the thesis as stage two of the social reorganisation of biology. Underpinning the whole thesis is the understanding that omic science is driven by a *systems biology* (SB) approach to twenty-first century biology. The realisation of this will constitute stage three.

Computational biologists are also using a similar model of scientific practice in order to map, trace and direct future scientific practice. However in using this developmental model, the organisation of scientific practice may turn messy when boundaries need to be permeated, re-aligned and re-ordered in the movement from post-genomic science to systems biology science. Consequently the specific aim of this research is to trace how two of these maturing research areas, *'proteomics'* and *'bioinformatics'*, are *emerging* and *stabilising* within stage two of the omic model, and to explore some of the social issues that are being reordered within their infrastructure. Drawing upon thirty-one interviews the research provides valuable insight into the social construction of post-genomic knowledge and adds to the growing literature in the field of *science and technology studies* (STS) by revealing how socially constructed knowledges are translated and transferred within and between newly created scientific communities. This is achieved through an examination of scientific identity, interdisciplinary expertise and community-based standardisation.

Table of Contents

List of Figures

List of Tables

CHAPTER ONE:
STUDYING SCIENCE - SCIENCE AND TECHNOLOGY STUDIES (STS)

PROLOGUE

Being based in the School of Social Sciences at Cardiff University, my peer group during the process of the doctorate has comprised social science, education and criminology Ph.D. students. Over the course of the four years, I have discussed my work both formally, in presentations, and informally, in social gatherings, with my contemporaries. However, when my peers initially hear about the focus of the research I often get the response: "oh don't tell me about your work, I don't understand science". This reaction has led to an increasing awareness of how inaccessible science seems to students from other disciplines. Whereas other areas of research might be readily translated across social worlds[1], natural science[2] has a way of demarcating *'insiders'* and *'outsiders'*. I have therefore made a conscious effort to make this research accessible to a wider audience; a skill that science itself, through public engagement policies, is always endeavouring to master. In line with the above ethos the introduction to this thesis has been written with the intention of being as accessible and comprehensible as possible for a science studies thesis.

INTRODUCTION

In this thesis I track, map and analyse how twenty-first century science is managed and organised. The primary aim is to look at the impact 'omic' science has had on biology as a *profession*, and as a form of *knowledge production.* Focussing on issues of scientific identity, research collaboration, disciplinary expertise and pedagogical routes, the thesis examines the ways in which two

[1] I am using the term 'social worlds' as described by Clarke (1991).
[2] Throughout the thesis I have used the term natural sciences to classify physics, chemistry and biology and to differentiate from the social sciences such as sociology or psychology. I realise that some authors separate the first category into natural and life sciences.

emerging research fields are attempting to secure coherence and stabilise into recognisable disciplines.

One significant development that has occurred in the wake of the Human Genome Project (HGP) has been the formation of new research areas. Two of these post-genomic research areas have been termed 'proteomics' and 'bioinformatics'. In line with the contemporary nature of new omic sciences these emerging research areas are inherently interdisciplinary, and are ordered in such a manner so as to produce multi-skilled researchers. The impact of this re-alignment has meant computer scientists, mathematicians, chemists and engineers have crossed traditional disciplinary boundaries and been welcomed as relatively new actors working in biological science.

This new style science is also a rich arena for studying the social practices involved in science since it is an area that truly reflects science and technology in motion. Concentrating on the task of ordering the complex web of new biological knowledge, this thesis highlights how new boundaries are being negotiated and renegotiated, constructed and reconstructed, and maps how post-genomic knowledge, objects, technologies and actors are translated across them. In essence, the thesis traces the development of proteomics and bioinformatics and analyses the ways in which these research areas are beginning to stabilise and solidify into recognisable and identifiable fields of research.

'SENSATIONAL' SCIENCE

> "The desire to complete the 'big picture' put forth by Newton, Darwin or Einstein has required the mass retraining of scientists in new techniques and methods, new ways of seeing the world, and sometimes the developments of new instruments of investigation. This reorientation, while invariably resisted by the scientific orthodoxy was at least financially tolerable. But as science has come to be so thoroughly involved in the economic and political maintenance of the societies housing its pursuit, any truly revolutionary project in science today would pose as great a threat to societal stability as a political revolution normally would" (Fuller 1997, p142).

To understand science we need to understand how social actors understand the world. In essence, we need to understand human socio-cultural interests. In this thesis I make sense of science by linking the way in which social actors understand the world, which I argue is through stories[3], with the way in which biologists are trying to comprehend developments in twenty-first century biology, or as Fuller (1997) phrases it: "...the[ir] desire to complete the big picture" (p142). Biographical accounts and oral justifications are also useful devices to translate knowledge across the social worlds of the scientists, with their specialised language, and the non-scientists, with their ubiquitous language. For example, Geesink, Prainsack and Franklin (2008) argue that in relation to stem cell research: "a good story to tell is crucial to fundraising for research, be it public or private, and for making the field acceptable to the public" (p1).

As far back as the ancient Greek natural philosophers whose interests were exploring matters of epistemology and logic, through to large-scale modern biological projects such as the Human Genome Project (HGP), the story that has connected all autonomous scientific disciplines with one another, and the story that has connected scientific theory to religious theory is the story of their ultimate goal. Each discipline or theoretical position, in one way or another, attempted and continues to attempt to understand why and how we are here by furthering our understanding of how life was created and helping to explain how it continues to exist. They are all accounts of comprehension in which each discipline endeavours to reveal some secret of life 'on earth' in order to help explain the 'essence' behind our existence. Individually, the disciplines want to place a sense of order on the particular parts of the world that they are detailing, whether it is biological, physical, chemical, psychological or social, and then explain those generated *'facts'* to the rest of the human race. Collectively they each want to justify their own perspectives, ideas and community identity.

[3] It is pertinent to emphasise at this juncture that this thesis is not a narrative approach to science and medicine in the tradition of Mulkay and Gilbert (1984), Williams (1984) or Frank (1995). Rather, I have drawn upon ideas of what stories convey to illustrate how the use of the dominant method of interviewing lends itself to a type of storytelling on behalf of participants. This produces an interviewees' retrospective of science in which they may expand and contract their accounts of the science they perform.

Since the mapping of the Human Genome, particular expectations are being relayed about the future of biomedical science. Such phrases as *'finding the holy grail'* and *'unlocking the code'* have been used to describe the aims, expectations and achievements within biology. This has been represented in literary texts such as: *The Code of Codes: Scientific and Social Issues in the Human Genome Project* (Kevles and Hood 1992); *The Book of Man: The Human Genome Project and the Quest to Discover Our Genetic Heritage* (Bodmer and McKie 1995) and *The Human Blueprint* (Shapiro 1992). Thus it would appear that this breakthrough in science has been met with great *'imaginations'* that finally biomedical scientists can make giant leaps in understanding why we are here and how we continue to exist. It is built on the premise, to use a tag line from the television show *the 'X-Files'*, that 'the truth is out there'. The truth being that biology is more than just a descriptive science and that there is an underlying digital code behind the biological mystery of life (Hood 2003). Moreover, this code can be unravelled and deciphered so that scientists can reveal an answer to how nature *really* works. It is assumed that all scientists need to do is crack the code and then reach a resolution.

Despite omic biology resurrecting, reintroducing and reinterpreting a number of the seventeenth and eighteenth century principles of modernity such as *truth*, *progress* and *scientific universality* (Chapter Three), discovering that there is a 'chest of treasure' at the end of the treasure map, or unearthing the 'holy grail', is epistemologically, scientifically and organisationally distinct from the process of working out how to open it, and then understanding its contents. As in the case of many pirate movie stories, once a box is discovered, characters need to learn how to unlock it. In many instances the discovered chest is shut or the final door closed, and so keys need to be cut and shaped to fit the locks in order to reveal the contents. This is also true in the biological world where discovering and sequencing a genetic code is one step but deciphering what it all means involves further research.

The genetic or genomic code is not an easy one. It should be no 'mystery' it has taken so many hundreds of years to reach today's level of understanding. Most of the codes formulated on the four nucleic acid bases, Adenine (A), Guanine (G),

Cytosine (C) and Thymine (T), are interwoven with one another. They have evolved over many thousands of years, and despite their startling level of conservation, the *right tools for the job* (Clarke and Fujimura 1992) are required to unlock them. It is believed in the biological world that one of these *right tools* is bioinformatics. This thesis argues that not only must the correct tools be used to cut, craft and shape the keys (bioinformatics), but the correct people must be found to deploy and manipulate the tools. As is described in Chapters Seven and Eight, these people are interdisciplinary researchers and cross-boundary demiurges[4].

Thus, if the dominant aspiration of new biological science is the quest to crack the code of life, then the dominant aspiration of those studying how science works is to examine how science positions and equips itself to crack the code, and how the revelation of the code is then interpreted by scientists and (re-)presented to the remainder of society. As Gary Alan Fine (2006) states: "the mission of science is to present the contours of the 'real world' in a way that audiences accept" (p12): making sense of the world in an ordered and understandable way and translating that story to those that science is intended to serve.

SENSING SCIENCE

The breakthroughs in science in the twentieth and twenty-first century have simultaneously affected and been affected by the transition from 'small science' to 'big science' (Price 1965). In biology this has meant that science today is not just new but it is also remarkably large in scope (Hevly 1992). It is bigger organisationally (in terms of the numbers of actors, countries and organisations involved), bigger mechanically (as a result of advances in bio-technologies), and bigger epistemologically (as some of its epistemological boundaries have been loosened). Biology's aspirations of what it can reveal have grown, and its outputs accelerated, yet it remains fascinating and bewildering to actors both *inside* and *outside* its community. One such *outsider* was the nineteenth century American satirist Mark Twain. On discovering that the Mississippi River had reduced in

[4] See Chapter Seven for an explanation of cross-boundary demiurges.

size, Twain summed up the fascination that we have with science by observing that:

> "There is something fascinating about science. One gets such wholesale returns of conjecture out of such a trifling investment in fact" (Mark Twain 1883, p173).

Today, with the development of high-throughput technologies that generate masses of raw data, Twain's comment can be turned on its head. Instead of getting such wholesale returns of conjecture out of such trifling investment in fact, it could be argued, that one gets such small returns in conjecture out of such huge investment in fact.

It is fair to state that science can no longer be assessed and studied without including the technologies that have been created to generate masses of data. This data is represented within biology as accomplished *facts*. Within the *social worlds* of genomics and post-genomics these technologies may include the physically imposing *mass spectrometer* (MS) or the comparatively much smaller and compact *microarray* technology. What they have in common however, is that they generate voluminous data on a scale never before seen in biology. The automation of technology has created a new sense of what biology can achieve, not least the possibility of mapping the genetic conundrum (the As, Cs, Ts and Gs) that nature has provided. The high-throughput technologies are able to extract biological information by metaphorically drilling down to the core of an organism's existence and elucidating these ciphers. This is not the Artificial Intelligence age however, but rather it may be better termed the Automated Information age, since in the current climate it could be argued that we do not have the wisdom to match the rapid development of technologies (Chapter Seven). It seems we might have the biological data (information) but we do not necessarily have the understanding (intelligence).

Yet if it is science's role to study nature, whose role is it to study science? Questions need to be asked by those *outside* of the scientific community to investigate how science has developed and how it has acquired its current position. Such questions may include: How closely does science represent

nature? Are the aspirations of the new biology realistic? In what ways are the identification of codes by technologies presented as scientific 'facts'? How are scientists making sense of such an abundance of 'facts'? Is there a discrepancy between data generation and data understanding?

The answer to the question of whose role it is to study science has often been left to scholars who have been described as 'historians' or 'philosophers' of science. The remit of the former has been to record the history of scientific development under the ethos that yesterday's biology is today's history, while the remit of the latter is to ask the philosophical question: what actually is science? In fact, John Ziman (1984) begins his book, '*An introduction to science studies: the philosophical and social aspects of science and technology*' by asking that same question, what is science? In his response, Ziman (1984) concedes that it is "much too grand a question to be answered in a few words" (p1). Nonetheless, he does suggest that the answer may lie in which part of science the questioner focuses on, and on what the same researcher is hoping to identify from the questions they ask. In stating this, Ziman recognises a comparatively new breed of researchers studying science: "...the history of science, the philosophy of science, the sociology of science, the psychology of creativity, the economics of research, and so on...", whom he states have different agendas to the philosopher or historian and so may ask different questions: "each of the metascientific disciplines...seems to concentrate upon a different aspect of the subject, often with quite different policy implications" (p1)[5].

This thesis may come under the classification of what Ziman (1984) presents as the sociology of science, or its more contemporary terms - science studies or science and technology studies (STS), which emphasise the fundamental role technologies are now playing within science. My focus is analysing creativity and expertise within biology, and investigating how the movement to omic biology has affected knowledge generation and transfer within different scientific sub-disciplines. Moreover, the specific interest is on interdisciplinary research

[5] This is a simple but important methodological point, because by asking different questions you will undoubtedly collect different answers. It is at this early juncture where the researcher begins to have an influence on their research project.

and the use of socio-bureaucratic devices such as the creation of community standards to facilitate, organise and make sense of emergent sciences. Therefore I am also interested in Ziman's (1984) question, what is science? And, in part, I have already begun to answer the question 'what is twenty-first century biology' during this introduction. More specifically, however, and perhaps more sociologically, I am interested in how omic biology has been able to manoeuvre itself into a position to create 'facts', the process by which those 'facts' are institutionally verified, the ways in which scientists overcome the obstacles and uncertainties involved in omic biology, and how knowledges and identities are created, debated and translated. These questions are the generic themes that run through a thesis focussing on an area of science which, some authors have suggested, has gone through a Kuhnian-type paradigm shift (Collins *et al.* 2003). To this end, it could be argued that big biology has (i) transformed the nature of work within science and with it changed science's relationship to funding, (ii) altered the dynamic between scientist and machine, and (iii) changed the meaning behind scientific claims on truth. New imaginations of what biology can achieve (unravelling the truth) are also beginning to be supported by scientific substance (new genetic information). Yet there is still the further question of how biologists are managing to make sense of this substance? Subsequently, the ultimate aim of this thesis is to examine how omic biology impacts upon the work of scientists and to discover some of the implications of this scientific reorganisation.

THESIS THEMES AND QUESTIONS

According to some authors biological science in the early part of the twenty-first century can be termed the post-genomic era (Campbell and Heyer 2003). In adopting this term they are not advocating the cessation of sequencing and analysing genomes (genomics), but rather the terminology signifies that some of the technical barriers preventing high-throughput biology are now being or have been resolved. This has meant biology has seen the emergence of new post-genomic research areas that have the potential to revolutionise health, medicine and scientific knowledge, which in turn will have an enormous impact on society. Two of these emerging post-genomic sciences are proteomics and bioinformatics:

two research fields at the heart of biology's new aspiration and two areas which have considerable research council funding. They are defined as:

Proteomics is the high-throughput science of identifying and analysing the full set of proteins produced by an organism during its life (Liebler 2002).

Bioinformatics is the interdisciplinary field of biology, computer science and applied mathematics. Bioinformatics' primary role in the new biology has been to make sense of large amounts of unorganised ratio data by reordering them into coloured clusters (Campbell and Heyer 2003).

In this thesis I focus on these cutting-edge interdisciplinary areas within biology, and assess how science is creating a stable scientific infrastructure in order to begin answering its own twenty-first century biological questions.

The development of new biological techniques such as bioinformatics and proteomics has led Atkinson and Glasner (2007) to observe that: "established ways of working as biologists or clinical scientists must be complemented by new skills and new inter-disciplinary teams" (p9). Others take this a little further and claim that old, traditional disciplinary boundaries are shifting and that new scientific paradigms are being constructed. This in turn is having a direct impact on the identities of scientists as new scientific infrastructures are created, and new skills and languages required (Evans, Plows and Welsh 2007). Issues surrounding new scientific identities and new types of interdisciplinary communication are explicit within the five empirical chapters of the thesis (Chapters Five to Nine), while this study follows a similar line of questioning to that of Evans *et al.* (2007) by asking the following five research questions:

(i) How do scientists claim an identity in a post-genomic era?
(ii) How do scientists make sense of emergent interdisciplinary research?
(iii) How have the research areas of bioinformatics and proteomics emerged and how are they beginning to stabilise?
(iv) How is the concept of expertise in new research fields constituted?
(v) How 'modern' is contemporary biology?

The focus therefore is on the importance of communication and collaboration in omic biology, and particularly on how specialised scientific knowledge is transferred between heterogeneous actors. In summary, the thesis evaluates how scientists go about cracking the 'code of life' that they are attempting to map, and provides data describing the way in which dry (computational) science is aiding more traditional wet (bench) science in this mammoth quest.

WHY STUDY SCIENCE?
NEW BIOLOGY'S IMPACT ON SOCIETY IN THE UK

The Department for Trade and Industry (DTI) states on its web-site that over the last ten years the UK government has substantially increased spending on science (DTI 2006)[6]. From 1997 to 2007 the overall science and technology budget has doubled. The web-site confirms that the budget for the year 2007-2008 will have risen to nearly £3.5 billion per year. As part of the increased expenditure, the UK government has increased funding to the Biotechnology and Biological Sciences Research Council (BBSRC), which along with the Medical Research Council (MRC) is the principal funding organisation for research across the biosciences in the UK. In turn the BBSRC will continue to promote multidisciplinary research by investing £25 million in Integrative and Systems Biology. This new investment will be used, in part, to enhance computational methods and bioinformatics across research in bioscience. It is clear from these statistics that science, and in particular biological science, is growing in the UK. There is also a suggestion that the way science is being structured is having an increased impact on society (Brenner 1998). This new emphasis and significance placed upon biological science is possibly reflected in the UK by the increase in expenditure on biological science supported by public taxes.

When discussing my research title with non-STS Ph.D. students they have often asked why a social scientist is working in a traditionally natural scientist's field. My response to this line of questioning is to state that if science is really for society and not just funded by society, as the statistics illustrate, then science and

[6] Although this may well change as a result of the current economic downturn in the UK.

society need to have a closer relationship. One way to do this is to open up the 'blackbox' of laboratory science in order to lay bare the concerns, issues, feelings and hard work conducted by cutting-edge scientists. It is from this position that I believe sociologists of science may analyse scientific practice from an alternative perspective to the alternative researchers that Ziman (1984) lists.

HOW TO STUDY SCIENCE: SCIENCE COMMUNITIES AND SCIENTIFIC PRACTICE

Collins (1992) states that within science:

> "the locus of knowledge is not the written word or symbol but the community of expert practitioners (this includes communities of theorists). Individuals' knowledge must be acquired by contact with the relevant community rather than by transferring programmes of instruction" (p159).

His statement is a strong voice for empirical social science. Collins' belief is that a community of knowledge cannot just be found within the production of texts that are disseminated in the public domain. Collins (1992) and his contemporaries (Knorr-Cetina 1999; Pickering 1984) argue that if someone wants to *do* science, they cannot do it by simply mimicking text books, since text book science does not reveal how science actually works. In other words, text book science does not present the inner workings of scientific practice; it just records the outer surface and provides a cleansed story. This view is shared by Mulkay (1976) who advocates the "close analysis of the development of research networks and of the social processes [particularly] by means of which standards of scientific adequacy and value are negotiated and applied to knowledge claims" (p639). For both Collins and Mulkay the starting point for any analysis of science is the scientific community; a social group whose members may share the same cultural, social or theoretical beliefs. The community is the locale where scientific knowledge is created, and is the point of departure in making sense of how science makes sense of scientific substance.

Science communities are constituted by relationships in which tacit, specialised and emergent knowledge is transferred and transmuted. Within the socially constructed spaces scientific practice is validated and knowledge transmission is

encouraged (Chapter Five). The scientific community is the setting where theoretical ideas are negotiated and where self-identity is institutionally verified and regulated. The boundary created by the community, whether physical as in the walls of a laboratory, or social as in a disciplinary boundary, establishes and legitimates the actions contained within it[7]. For example, scientists can become experts in a particular domain and can perform valid scientific experiments within that space, which if performed elsewhere may look like an absurd cultural ritual.

When small scientific communities (scientific laboratory groups for example) are combined and linked together they can create a larger scientific community, as for example the proteomics community. These communities network together through communication and collaboration to create the structural framework of science. In particular the scientific networks reinforce the scientific organisation as the dominant ideology within society by creating more disciplinary *facts* that can be used as evidence for science's accuracy and rationality. It is no coincidence then that STS scholars have used these settings to study local groups in action (Collins 2004a; Knorr-Cetina 1999; Latour and Woolgar 1979; Law 1994; Traweek 1988). They have done so in order to explore, in detail, how the interaction at this intermediate level affects the wider global structure. Their assumption is that the universal is often contained in the particular and the particular lies at the community level. The focus of this thesis is similar, although arguably not as straightforward. The objective is to study both physical groups and communities such as the European Bioinformatics Institute (EBI), but also wider physical and virtual networks such as proteomics and bioinformatics communities, which in the case of this research are spread across the whole of the United Kingdom (UK). Consequently I argue that the term 'community', building on Kadushin's theory of the social circle effect (Kadushin 1966, 1968), can have more than one level of definition. In the proteomics social world, there is the laboratory or departmental community level (the EBI) with direct ties between members, and also a larger informal networked community based on areas of interest and affinity (affiliated scientists).

[7] But obviously not scientific malpractices such as the Hwang Woo-Suk stem cell affair (see Rincon 2006).

The focus in this thesis is predominantly on this second tier of communities situated at the network (interest) level; a kind of disciplinary network or epistemic community (Knorr-Cetina 1999). Nevertheless the research is still performed by concentrating on local areas within the UK. In accordance with this, the study researches how processes of standardisation, communication, expertise, knowledge-transfer and boundary formation are changeable and social activities negotiated at the local level, but proceed to have a large impact on the wider structural framework. This is achieved by studying communities.

THESIS SYNPOSIS

The thesis consists of ten chapters including five analytical chapters. The short descriptions of these chapters that follow are intended to act as navigational guides for the reader by offering a synopsis of their contents. It is hoped that this compendia will assist the reader in understanding the research aims and research questions illustrated earlier and also the rationalisations behind these choices.

Chapter Two: 'Proteomics and Bioinformatics – A Social Scientists' Primer'
The aim of this chapter is to familiarise the reader with the historical developments in both bioinformatics and proteomics. It is written in a style intended to provide an accessible understanding of the technical and scientific detail. This forms part one of two literature review chapters.

Chapter Three: 'Standards, Boundary Classifications and Paradigm Shifts'
This chapter is the second of the two literature review chapters and focuses on scientific standards, scientific methods and scientific collaborations. The first part of the chapter highlights the way in which standardisation is a major component of scientific communication, scientific collaboration and scientific stabilisation and presents three short stories (the Linnaean taxonomy model, the UK drug classification system and the QWERTY keyboard design) as examples of this. The second section of the chapter positions omic biology within the literature of Kuhn's (1970) scientific paradigms.

Chapter Four: 'Methodological Reflections: A Social Scientist in a Natural Scientist' Setting'

In this chapter I reflect upon the experiences of performing qualitative fieldwork in a scientific setting. It pays particular attention to the specific issues of Interviewing Elites and Interactional Expertise. The idea of Action Research is also discussed from the position that I found myself inadvertently changing the scientific setting I was studying. By presenting an account of personal experiences and particular incidents that occurred during the fieldwork, the aim of the chapter is to show how Qualitative Inquiry (QI) is a flexible process resulting from negotiation between the researcher and the research participants.

Chapter Five: 'Beyond Boundaries: Performing the Promise of Proteomics'

Chapter Five is the first of the five data chapters. The chapter presents proteomics as a buzz word, and analyses it as *'proto-boundary object'* in a period of scientific stabilisation that I call *phase* zero. The chapter also focuses on science's relationship to funding and demonstrates how the new research area of proteomics is able to mobilise disparate scientific actors because of the hype and promise invested in the term. In recognising that science is a social world, the chapter argues that a proto-boundary object may fail to stabilise into a more robust boundary object and illustrates how scientific research areas continually attempt to re-brand themselves.

Chapter Six: 'Scripting the Gold-standard: Whose Standard is it Anyway?'

In this chapter I focus on an organisation called the Proteomics Standards Initiative (PSI). Their remit is to construct community-based data reporting standards for the proteomics community. I track the development of these standards as devices to improve communication between actors within the proteomics community and also between other post-genomic communities. I also illustrate that a secondary function of the standard is to help identify and authenticate proteomics as a maturing research area. In this sense, standards in proteomics, which I represent as being driven by a particular scientific future, are helping to stabilise the research field by transforming it from a proto-boundary object towards a more robust boundary object.

Chapter Seven: 'Computing Biological Identities'

The fundamental research question of this chapter is to discover whether bioinformatics is a service to biology or a freestanding discipline in its own right. As such the chapter explores ideas of creativity in science, and separates the ideas of information generation from knowledge creation. Specifically it assesses whether the technical use of technologies is being recognised as a creative task or not, while also highlighting how bioinformaticians believe that their profession is having trouble claiming its rightful identity in twenty-first century biology. The chapter also illustrates how some bioinformaticians separate their own discipline into *bioinformaticians* and *bioinformaticists* as ways of highlighting their creative input.

Chapter Eight: 'Matchmakers and Speed Daters: Cross-collaborative arrangements in bioinformatics and proteomics'

Bioinformatics and proteomics are highly interdisciplinary fields in which biologists, chemists, mathematicians and computer scientists find themselves working together. In this chapter I present some of the different languages that are involved in post-genomic science. I then present five communicative and collaborative mechanisms that *matchmakers* utilise to aid communication and comprehension within emergent fields. This is during a period that I tentatively call *permodern science*.

Chapter Nine: 'Educating New Chameleon Scientists'

Chapter Nine is the final data chapter. The chapter concentrates on craft knowledge-transfer within science and on learning within cutting-edge interdisciplinary research areas. It focuses on how bioinformatics may develop, the ways in which interdisciplinary research in post-genomic science is taught, and how experts become trained in a new research field. In this regard the chapter portrays the construction of academic degree courses as forms of scientific stabilisation.

Chapter Ten: 'New Disciplines: Emergence and Stabilisation'

In the concluding chapter I comment and reflect on the findings of the research. In particular, I discuss the status of the arguments as they have been developed

during the thesis and integrate the conclusions made in Chapters Five to Nine. Finally, the chapter concludes by proposing some questions that require further research.

CHAPTER TWO:
PROTEOMICS AND BIOINFORMATICS:
A SOCIAL SCIENTIST'S PRIMER

PART ONE - PROTEOMICS

INTRODUCTION

The purpose of this chapter is to introduce the reader to two emerging and inter-dependent post-genomic scientific disciplines, 'proteomics' and 'bioinformatics'. Written by a social scientist and with a social science audience in mind, the chapter follows the lead of other works in science and technology studies (STS) (Law 1994; Rabinow 1996) by explaining some of the basic principles of molecular and computational biology, without delving into deep technical detail. The initial focus is to explain the function and structure of proteins, and the level of description portrayed there sets the tone for the level of detail the reader can expect in the remainder of the chapter. Following Crane (1972), it is understood that 'the sociologist' must at least be able to engage with the technical ideas produced by the intellectual subject that they are studying:

> "It is not surprising that the sociological analysis of the production of science, ideology, philosophy, religion and the arts has been largely neglected since few sociological problems are so complex as that of understanding the social institutions that produce ideas. In dealing with these types of phenomena, the sociologist is faced with the problem of not only understanding the social relationships between individuals but also of understanding the ideas themselves, which can be highly technical and abstruse. Even if the sociologist elects not to become an expert on the details of his subjects' intellectual productions, he cannot ignore the nature of these activities entirely since presumably they affect in some way the social relationships among his subjects, and the latter in turn affect the production of ideas" (p2).

This chapter also acquaints the reader with the relevant knowledge required to understand the scientific issues and terms embedded within the thesis. By providing this, the chapter illustrates some of the complexity found in cutting-

edge science, highlights the need for a functioning, efficient and collaborative scientific network to make sense of this intricacy, and portrays all the convoluted stories of scientific, technological and informatic development that create the present 'post genomic' era coined by, among others, Blackstock and Mann (2001).

In summary, the chapter provides brief accounts and histories[8] of scientific developments in the worlds of 'proteomics' and 'bioinformatics' as a way of recording their current level of stabilisation. The purpose behind this is to help the reader better understand the worlds of the actors studied. The period of history explored in the chapter begins in the early part of the twentieth century, but is particularly focused on the period from the 1950s to present day scientific activity.

WHAT ARE PROTEINS?

There are far more proteins than genes in the human body. In August 2005, it was calculated with confidence that there are 22,118 genes in the human body (see McNally and Glasner 2007). In comparison, it is estimated that humans could contain anywhere between 150,000 to over 1,000,000 proteins (Twyman 2004), the substantial discrepancy highlighting the difficulty and complexity in recording an accurate count[9]. But what exactly are proteins and what do they do?

When most people consider proteins they possibly think of the protein content in foods or perhaps protein shake supplements used to increase muscle strength. In fact, proteins are a type of intricate class of molecule called polypeptides. Polypeptides are made up of thousands of tiny units called amino acids created in a "condensation reaction between the amino acid group of one amino acid and the carboxyl group of the next" (Strachan and Read 1999, p2). The resulting creations (the proteins) perform numerous essential functions within the body. These include acting as an enzyme to control the reactions within cells and helping to repair and replace human tissue. Proteins are so central to human

[8] Presumably derived from the words his and stories.
[9] Of the total number of proteins in the human body approximately 20,000 to 25,000 are thought to be encoded by genes (Liew *et al.* 2006).

biology that they constitute approximately twenty percent of the mass of an average person (Harwood 2002). Furthermore, they are the critical ingredients for most of the 'good' and 'bad' things that happen in the body. On the one hand, proteins perform most of life's indispensable functions by behaving as antibodies and hormones, while on the other - most diseases in the human body manifest themselves at the protein level of activity. Craig Venter, president of Celera Genomics, highlighted the importance of proteins in the body when he proclaimed that: "most biology happens at the protein level and not the DNA level" (as cited in Dennis and Gallagher 2001, p19). Proteins then are the ultimate performers in both 'health' and 'sickness' and are responsible for most of the cellular function in organisms.

Historically, the complex structure behind the make-up of proteins made attempting to understand proteins extremely troublesome. The hidden helix and multiple mosaic structure behind protein chains (their formation) meant proteins were not well understood well until relatively recently. A possible reason for this is that chains of proteins are synthesised by a linear sequence of approximately twenty amino acids (Frauenfelder 2002). Indeed, the structure of proteins is so dependent on the complex string of amino acids that amino acids are metaphorically known as the 'building blocks' of protein structure. The result of this intricate chain means that the first complete sequence of a peptide (a small chain), namely the hormone insulin, was not accurately annotated until the mid-1950s (Ryle *et al.* 1955). Furthermore, prior to 1945, as Attwood and Parry-Smith (1999) explain, there was not a "single quantitative analysis available for any one protein" (p1.), while it was another five years before the first enzyme, 'ribonuclease' was completely sequenced in 1960 (Hirs *et al.* 1960). Although some areas of science move extraordinarily quickly, understanding proteins is taking a comparatively long time.

A decade later, and throughout most of the 1970s, it was widely believed within the scientific community that a single gene was responsible for one polypeptide and that the human body contained more genes than proteins (Lewontin 1992). But in the twenty-first century, and as a consequence of increasingly developed automated technologies, it is now recognised that there are indeed many more

proteins than genes. In the mid-1990s researchers calculated that on average one gene could code from between three and twenty protein spots, many of which being modified polypeptide chains (Wilkins *et al.* 1996a). The idea of modification and mutation in the amino acid sequence, combined with the multifaceted structural make-up of the chain is the key to understanding the complexity of protein structure, and perhaps the reason why proteins have been so poorly understood to date (Lesk 2002).

COMPLEX AND CONVOLUTED CHAINS

According to Attwood and Parry-Smith (1999) there are five different levels of protein composition. This structure is central to how proteins function. The secondary and tertiary levels of the structure refer to folds within the protein, which determine its complex final three-dimensional (3-D) shape. The convoluted 3-D shape intensifies the thousands of different functions that a protein may perform. The intricacy of protein folding is such that it has led Groβ (1998) to state that: "protein folding appears to be almost too complex for a complete description or for accurate structure prediction from sequence data" (pR308). Thus, many scientists agree that proteins are much more complicated than nucleic acids, illustrated by the fact that only sketchy estimates of the exact number of proteins in the human proteome exist.

To make matters more complicated, proteins are also prone to modification. This is partly due to the fact that messenger RNA (mRNA), the central copy of DNA, which is used as an original pattern when a cell creates a protein, can undergo a series of edits after it is originally copied. The result is that a number of different proteins can be created from the same gene and these proteins can appear in a number of different guises. Even after the translation from mRNA to DNA, a protein can undergo a number of further transformations. So while biochemists, before the advent of large-scale biological projects and improved technologies, made quite accurate educated guesses about the number of genes in the human body (approximately 20-30,000), estimates of the number of proteins proved to be less precise. In essence, the relationship between sequence and structure in proteins is such a challenging problem for biochemists and informaticians alike,

that it has led some to believe that "the world of individual proteins is...far larger, more complex and potentially more rewarding than the world of the genome" (Anderson, Matheson and Anderson. 2001, p4). The story of protein (under-) development during the 1950s to 1970s helps therefore to explain how the complexity of protein structure comprehension is still a major challenge for scientists in twenty-first century omic biology.

WHAT IS A PROTEOME? FROM PROTEINS TO THE PROTEOME

If we move on thirty years, leaving the ambiguous protein research of the 1970s behind, one of the burning question in molecular biology today is: 'What is the proteome'? According to Service (2001), if you were to "ask a dozen people that question you might get a dozen different answers" (p2074). Garavelli (2002) even goes as far as to suggest that attempting to *define* a proteome is a challenge within itself for scientists, let alone *making sense* of one. Nevertheless, it seems that most specialist definitions concur to characterise a proteome as an organism's complete complement of all proteins in a cell or the "protein complement of the genome" (Liebler 2002, p3). Whereas a protein is an individual polypeptide, the proteome is the interlinking total number of proteins in any one organism.

The invention of the term 'proteome' is credited to the Australian postdoctoral fellow, Marc Wilkins, now a lecturer at the University of New South Wales (UNSW) and co-founder of the company Proteome Systems in the 1990s (Wilkins *et al.* 1996b). It is reported that Wilkins got tired of having to repeatedly write the sentence "all proteins expressed by a genome, cell or tissue" (Cohen 2001, p56) when writing a scientific paper to support his Ph.D. thesis. After rejecting his initial choice of *'proteinome'*, he replaced the sentence with his new word *'proteome'* meaning "the total set of proteins expressed in a given cell at a given time" (Dove 1999, p233). In 1994, during a meeting of the two dimensional polyacrylamide gel electrophoresis (2DPAGE) conference held in Siena, Italy, Wilkins publicly used the word *proteome* for the first time to describe the "protein complement of a genome" (Anderson and Anderson 1982). The term has since seemingly prospered, unfurling through the protein chemistry and molecular

biology communities to attract large amounts of money. This relationship between scientific funding and proteomics is explored in Chapter Five.

PROTEOMICS: A SUCCINCT DEFINITION AND HISTORY

The word *protein* originates from the Greek word 'proteos' meaning primary (McNally and Glasner 2007) or 'of first rank' (Strachan and Read 1999) and is derived from the Greek sea God, '*Proteus*', son of '*Poseidon*' (Graves 1955). From the term *proteus* came the adjective *protean* meaning flexible and capable of taking many forms. This meaning explains why Mulder first suggested the name '*protein*' in 1838 to describe this highly varied and complex- shaped organic molecule (Stent 1971).

The methodological study of the proteome is called *proteomics*. Its etymology derives from the prefix '*prot*' from *protein* and the ending '*omic*' from *genomics*. Proteomics is used to describe the identification, analysis and quantification of large amounts of proteins, almost always with the aid of computers (Pandey and Mann 2000). Anderson and Anderson (1998), stalwarts in the field of protein research, define proteomics as: "the science that uses quantitative protein level measurement of gene expression to characterise biological processes and decipher the mechanisms of gene expression control" (p1853). Essentially, proteomics is the global scientific study of the multiprotein system (Hood 2003; Liebler 2002).

In comparison to other contemporary scientific fields such as genomics and metabolomics, proteomics has developed quite slowly though. The discipline, if not the actual word, can be traced back thirty to forty years to the late-1970s and before ideas of the Human Genome Project (HGP) had even been discussed (McNally 2005, McNally and Glasner 2007). To this end it benefited from two rather different techniques: *mass spectrometry* and *electrophoresis*. A short history of these techniques and an illustration of their impact on proteomics development is outlined below.

ELECTROPHORESIS, MASS SPECTROMETRY AND PROTEOMICS

Electrophoresis is a technique that uses electric currents to separate mixtures with varying surface charges. The method can be traced back to the 1930s when Swedish scientist Arne Tiselius (1902-1971) developed a tool for separating proteins in solution (Tiselius 1937). Tiselius was later acknowledged for his efforts in this field with the Nobel Laureate in Chemistry in 1948. Nonetheless, it was not until the 1970s and early-1980s, after the mass production of electrophoresis machines that this specialised technique became widespread. After initial publications on the isoelectric method[10] by Klose and Spielman (1975), O'Farrell (1975), and Scheele (1975), new technologies were developed that built around two-dimensional electrophoresis (2DE). This technique enabled human proteins to be separated in a gel, tracked, mapped and then analysed on a substantial scale by applying the technique simultaneously in two opposite directions. In one dimension the proteins can be separated by molecular weight and in the other dimension by charge (Liotta and Petricoin 2000). The result leaves a variegated and intricate map on a gel, often with over one thousand scattered spots, each one representing an individual protein (Wade 1981). Although this technique is slow in comparison to many high-throughput automated technologies in big biology today, the capability of the electrophoresis technique in isolating and identifying proteins explains the early strong bind that still exists between proteomics and electrophoresis into the early twenty-first century.

Mass spectrometry (MS), on the other hand, is a technique used for measuring the molecular mass of ions. Using soft ionisation methods (by adding enough energy to the molecule) it allows proteins and peptides to fly through a spectrometer. The fragments are then analysed based on their attributions including their sensitivity, their mass range and their charge.

The development of MS can be traced back to the work carried out by Joseph John Thomson (1899) and his protégé Francis William Ashton (1920) at

[10] A technique used to separate different molecules by their electric charge.

Cambridge University in the nineteenth century. Ashton used the mass spectrograph to discover a number of isotopes in non-radioactive elements. His work was an extension of Thompson, who developed the first mass spectrometer when investigating the conductivity of gases. Both men were awarded Nobel prizes for their work, Thompson in 1906 in Physics, and Ashton in 1922 in Chemistry.

However, it was not until the 1950s and the development of the gas chromatography mass spectrometer (GC/MS) that the origins of the modern MS emerged (Gohlke 1959). This technique was based on the coupling of the two instruments – GC and MS - to produce the prototype of many of the mass spectrometers in use today. Over the next thirty to forty years new developments in MS have seemingly appeared every few years. The primary types associated with the development of proteomics are the Matrix Assisted Laser Desorption/Ionisation Time-Of-Flight Mass Spectrometer (MALDI-TOF/MS) devised by Franz Hillenkamp and Michael Karas (1991), and the Electrospray Ionization Tandem Mass Spectrometry (ESI/MS) technique devised by John B Fenn (1989).

Despite both MS and GE having long and distinguished histories it was not until the early 1990s, a few years before Wilkins coined the term *proteomics*, that a breakthrough in proteomics activities occurred. In 1993, Henzel and colleagues combined the MALDI-TOF mass spectrometry technique with the modified 2-D gel electrophoresis technique to study proteins. Ever since this initial idea, protein mass mapping fingerprinting technique has been an indispensable tool used in proteomics research.

Today technologies are continuing to develop with the introduction of the Tandem Mass Spectrometer Peptide Sequencing with Nano-Electrospray Qtof (quadruple Tof) – MS/MS. On top of the analysis by the MALDI, this type of technology allows peptides to be analysed individually in a liquid rather than in a solid state. The development of this type of improved automation in technologies is removing many of the technical and scientific barriers that prevented proteomics prospering in the 1980s. Below, I describe some of these barriers and

illustrate how close the field of proteomics came to emerging in the 1970s and 1980s.

THE HUMAN PROTEIN INDEX (HPI)

It was on the strength of the two dimensional electrophoresis (2DE) techniques profiling proteins that initial murmurings of constructing large-scale protein databases emerged within the discipline of protein chemistry (Patterson and Aebersold 2003). In 1980, the Human Protein Index (HPI) task force was initiated following a review of the uses of two-dimensional electrophoresis (2DE). The group was asked to produce a human database of proteins[11] providing a wealth of information about each individual protein (Anderson and Anderson 1982). They proposed that each protein should be accompanied by a description listing the name of each protein, attaching any credible literature references on the function of the particular protein, and affixing a molecular map illustrating the encoded gene and the protein's corresponding amino acid sequence (Anderson *et al.* 2001). The result was an associated report (Anderson *et al.* 1980) recommending a single central protein laboratory to maintain the standardisation and verification of all protein data generation. The empirical work in Chapter Six will examine the construction of community based proteomics standards.

Regrettably for those behind the HPI task force, the election Reagan administration in the US marked a shift away from large-scale and expensive big scientific due to lack of funding. Without such support, work on mapping the Human Proteome was effectively suspended for over twenty years, until the idea was revived by Marc Wilkins' paper at the 1994 Siena meeting. The two major barriers therefore that prevented proteomics from flourishing in the 1980s were the complexity of protein structure (a technical issue)[12] and the lack of financial resources (a political issue). This example illustrates the complex relationships

[11] This was one of the first allusions towards big biology; big biology being a branch of big science, a phrase used initially in large-scale physics and chemistry projects funded by governments often during periods of wartime. For example see the Manhattan Project (Price 1965).

[12] Gupta and Guglani (2001) even suggest that if proteins did not have such intricate structures an in-depth proteomics database could have been set up long before the Human Genome Project (HGP) was devised.

science has with politics, society and culture (Jasanoff 2004). Focussing on issues of funding and technical barriers Chapters Five to Nine will consider how proteomics is stabilising in the 21st century.

BIG SCIENCE PROJECTS: FROM REDUCTIONISM TO HOLISM

Protein chemistry and protein studies are examples of two scientific disciplines that clearly represented the *reductionist* approach to science within biology during the 1980s. In respect to protein chemistry, Patterson and Aebersold (2003) argue that the discipline provided "the link between the observed activity of a biochemically isolated protein and the gene that encoded it" (p311). This method was then the epitome of reductionism as it entailed dissecting proteins on a small scale into their four or five levels of structural organisation.

The technological advancements in the 1980s in both computer and laboratory technology however, were an indicator of the changing tide: there was a movement from reductionist approaches in biology to large-scale sequencing projects. This development occurred alongside significant events such as; the launch of the first Compact Disc, the National Science Foundation (NSF) network linking up every university in the US, the emergence of Polymerase Chain Reaction (PCR) - a technique which allows copies of DNA to be reproduced quickly and easily (Rabinow 1996), and the production of the first automated DNA sequencer (Smith *et al.* 1986). The difference between the two approaches is subtly summed up by Fujimura (2005) who states that "in contrast to reductionist genetics [and protein chemistry], one could argue that systems biology is attempting to model biological complexities as organised systems in order to understand them" (p198). Therefore, even though technical and political barriers were apparent in the 1980s, there was still a gradual movement from studying individual proteins in isolation towards a more systemic view of proteins.

THE BEGINNINGS OF SYSTEMS BIOLOGY

The possibilities of an holistic approach to biology in the 1980s were certainly visible to Leroy Hood, now of the Institute of Systems Biology (ISB). Convinced of its emergence, he coined the term systems biology (SB). The term is defined today as:

> "...a study of biological system by the systemic and quantitative analysis of all the components that constitutes the system." (Patterson and Aebersold 2003, p312).

The fundamental aim behind Hood's systems biology theory is to provide an understanding of human physiology by comprehending how molecules (genes and proteins) interact in the global make-up of cells. Systems biology thus promotes the investigation of an organism and its interconnected parts as one system, rather than studying each individual element of that organism in isolation. Its underlying ethos is to use the functionalist organic approach: the whole is greater than the sum of its parts.

Although Hood began discussing macro approaches to biology in the late-1980s, (Moody 2004), this was not the first time a systems approach had been discussed. A number of leading researchers interested in a general systems approach to living matter also emerged in the 1960s (Mesarovic 1968). The leap from a systems approach in physics to a systems approach in biology was however met with criticism (Rosen 1978). Biology was pitched as an exceptional case within science and large scale projects were seen as redundant because the discipline was thought to be descriptive rather than predictive[13]. Despite this, further breakthroughs in robotics and informatics in the 1980s, allied with Hood's fierce commitment to the SB concept, meant that the foundations were laid for the emergence of an interlinked biological 'information age'[14]. In Chapters Six and Seven I explore the impact SB is having on the stabilisation and direction of proteomics and bioinformatics.

---

[13] Today, Hood (2003) believes biology has the advantage over sciences such as physics and chemistry since at its core is the genome; a knowable analogous programme.

[14] In her work, Keller (2000) explains how molecular biologists used information as a metaphor for biological sequencing.

PREPARING FOR THE HUMAN GENOME PROJECT

Biology's transition from a descriptive science to an informational science cannot be discussed without referring to the genome. The genome is the complete DNA sequence or genetic constitution of an organism. The word was created when Hans Winkler fused together the words *GEN*es and chromos*OME*s in the 1920s to make the new term *genome*[15]. At the substructure of the genome is the electronic code described in Chapter One, which is believed to contain all the vital information necessary to initiate and understand human development and physiological responses (Auffray *et al.* 2003).

During the same time as Hood began toying with the idea of a systems approach to biology, another development in the rhetoric of scientific delineation occurred. In 1986, sixty years after Winkler had invented the word *'genome'*, Thomas Roderick proposed the term *'genomics'*. Roderick used the word to refer to the scientific discipline of mapping, sequencing and analysing genomes when publishing his first editorial of a journal with the same name. He suggested that genomics should be recognised by the scientific community as an independent discipline rather than a small part of an already existing field. Although, autonomous recognition initially met with resistance, a consensus emerged in the late-1980s and early-1990s that genomics needed to play a vital role if systems biology was to be successful.

At the same time as the first copies of the journal *'Genomics'* were circulated within the scientific community, discussions were being held on how to conduct a worldwide sequencing project to map the Human Genome. In May 1985, a small number of scientists in California first put forward serious proposals to sequence the Human Genome; an endeavour that was thought to be one of the most challenging to ever face scientists and an enterprise that was mocked in the 1960s and 1970s (Watson 1990). A year later, in March 1986, Nobel Laureate Renato Dulbecco wrote a hugely influential article highlighting the potential of whole genome sequencing within Cancer research (Dulbecco 1986). He emphasised that

[15] Although, it should be noted here, there are alternative interpretations of the true source of the 'ome' suffixed word, for example 'biome' (Mennella 2003, McNally and Glasner 2006).

Human Genome was of paramount importance if science was to understand human physiology, declaring that scientific tools and technologies such as developments in DNA sequencing and cloning precipitated a new approach to cancer research.

Later in the same year, the National Academy of Sciences (NAS) set up a special National Research Centre committee in the United States. The committee was chaired by Bruce Alberts, an opponent and detractor of big biology (Alberts *et al.* 1988). Yet, even with Alberts on board, after fourteen months of deliberation, the committee concluded that in the interests of the development of global science a Human Genome Project (HGP) should be initiated. The future direction of the HGP was thus largely determined by this initial report (Olson 1993). Aspects of the movement to big biological projects and global networks are discussed in Chapters Seven and Eight.

THE HUMAN GENOME PROJECT AND THE HUMAN GENOME ORGANISATION

In 1989, the Human Genome Organisation (HUGO) was finally set up to promote international collaboration on the HGP (HUGO 2005). It was hoped that this organisation would administer and regulate the project in an open and effective way by bringing together some of the world's super science powers including the US, the UK, Japan, France and the USSR. In all, there were members from twenty three different countries from across the globe (McKusick 1989). Echoing a number of the aims set out in the original HPI, HUGO was founded:

(i) to determine the sequence of the three billion chemical base pairs that make up DNA,
(ii) to store the information in large databases,
(iii) to improve tools for data analysis and manipulation,
(iv) to transfer the related technologies into the private sector, and
(v) to address any ethical, legal and social issues (ELSI) that may arise during the HGP (HUGO 2005).

Finally, on October the 1st 1990, the HGP, in theUS at least, was officially launched with federal, private and public funding[16]. The intention was that an accurate classification of the euchromatic portion of the Human Genome be sequenced within fifteen years.

In 1995, five years after the commencement of the project, Collins (1995) stated that the project was "ahead of schedule and under budget" (p10821). The original timescale targets were being surpassed due to rapid improvements in computer technologies and laboratory methodologies during the 1990s. A clear example of these developments included a massive 1000 to 2000-fold increase in Smith and Hood's original DNA throughput sequencer invented in the 1980s (Hood 2003). This development led Collins (1995) to proclaim that:

> "this will only be the end of the beginning of the era of sequenced based biology, and continuous improvements in capabilities for large scale sequence analysis, placing megabase sequencing in the hands of an average laboratory, are to be expected" (p10822).

However just as things started to quieten down on the HGP front, Craig Venter established a new private organisation in May 1998. Building on his company *'TIGR'*, Venter and his colleagues established a new private organisation *'Celera'* and challenged the public HUGO to a race to completion. With equipment supplied by *'Applied Biosystems'*, Venter began using a new technology in genomics sequence mapping called 'whole genome shotgun'; a more rapid technique than that used by his HUGO counterparts. This challenge was perceived as a threat to the public HUGO consortium and the Wellcome Trust reacted by doubling their funds for the Sanger Centre research in the UK. Congress in the US also increased their funding. The result was a draft publication of the Human Genome produced by the international consortium in the journal *'Nature'* in 2001 (Lander *et al.* 2001), while Venter simultaneously published his version in the rival scientific journal *'Science'* (Venter *et al.* 2001)[17].

---

[16] In the US the project was funded by the National Institute of Health (NIH), the Department of Energy (DOE), and the National Human Genome Research Institute (NHGRI) (see McKusick 1997). Other large global financial support included significant funding from the Wellcome Trust at the Sanger Centre at Cambridge in the United Kingdom (UK).

[17] For 'good' political reasons the genome race was diplomatically declared a draw (Aach *et al.* 2001).

In retrospect, the challenge from Venter not only assured that the project was completed ahead of schedule, but also have increased the development, pace and quality of high throughput automated technology, while indirectly effecting the scientific future of proteomics[18]. Nonetheless the race between Venter and the Human Genome Consortium highlighted the large social interests and social relations implicated in big science projects. Factions and social interests within big science are examined in Chapter Five.

The draft publication was greeted with worldwide acclaim. In the UK, Prime Minister Tony Blair declared that:

> "Let us be in no doubt that what we are witnessing today is a revolution in medical science whose implications far surpass even the discovery of antibiotics, the first great technological triumph of the twenty-first century" (Watson, as cited in Dennis and Gallagher eds. 2001, p11).

While, US President, Bill Clinton stated that:

> "Without a doubt this is the most important, most wondrous map ever produced by mankind" (as cited in Cohen 2001, p55).

The full genome sequence was completed in 2003, two years ahead of the original predicted schedule. It was the first example of detection and discovery science in biology where all the rudiments of a biological organism were described and classified into an annotated database (Auffray *et al.* 2003*)*. Naturally, the availability of fully mapped genomes of particular organisms, coupled with advances in electro-mediated devices such as the World Wide Web (WWW) and the Internet, made Hood's initial visions of systems biology twenty years earlier more plausible. As Fujimura (2005) writes: "systems biology developed in positive response to the vast territories of information produced by the genome sequencing projects" (p195). While, as a way of illustrating its impact, Hood and his colleagues have recently published articles attempting to construct mathematical models depicting the structure of the organism under the premise of

[18] By changing the course of scientific history and impacting on the development of high-throughput technology.

a system's biology approach (Ideker, Galitski and Hood 2001). It seems unlikely this endeavour would have occurred without the HGP.

FROM THE PROTO-OMIC TO PROTEOMICS

Once the Human Genome had been sequenced, proteomics was seen as the next logical step in the omic concatenation (Cohen 2001; Tyers and Mann 2003). Proteomics is often referred to as a post-genomic science (Chapters Five and Six); others even believe it to be a paradigmatic shift in science (Boguski and McIntosh 2003). Unlike genomics however, which attempts to identify and sequence previously unknown genes, the ultimate aim of proteomics is to assemble a complete library of all proteins (Liotta and Petricoin 2000). As explained earlier, this is a much more difficult task because of the proteome's more complex structure.

Although the antecedents of proteomics can be traced back thirty years to Margaret Dayhoff's work on protein mapping (Dayhoff *et al.* 1965; Dayhoff and Eck 1970) and the Andersons' (1982) work on the Human Protein Index (HPI), it was genomics and, in particular, the Human Genome Project that triggered its current progression. Just as genes were the blueprint for proteins (Cohen 2001), genomics provided the blueprint for proteomics (Tyers and Mann 2003). The data and technology created by sequencing the genome benefited existing fields of research, of which proteomics was one. The rapid developments in genomic technologies could now be used to develop upon the Andersons' and Dayhoff's introductory work and begin to sequence the 'Human Proteome'. The additional access to databases containing sequenced genes could also hold the secret to the protein that it encodes (Patterson and Aebersold 2003). Notwithstanding these advances, in 2001, Cohen maintained that sequencing technologies used to separate and identify intricate proteins needed to improve rapidly because they were "still cumbersome and insensitive in relation to where they need[ed] to be" (p55).

GENOME TO PROTEOME: A PROBLEM AND A PROMISE

Whereas the genome is arguably fundamentally static, the proteome is dynamic (McIntyre 2005). Unlike the number of genes, the number of proteins in the human body changes throughout a person's life. This fluidity in their make-up has led many people to proclaim that there is no such thing as 'the human proteome', since the number and type of proteins not only differ significantly between individuals but can also differ within the same individual over time. This dynamic feature has meant the study of proteomics is both technically difficult and conceptually problematic. As a consequence, some scientists believe that taking a systemic approach to proteins is too complex and the idea of a Human Proteome Project (HPP) replicating the Human Genome Project (HGP) was often dismissed:

> "I just don't know when you'd ever say you finished. It's bad enough trying to figure out if you've finished the Human Genome project" (Scott Patterson, Celera, as cited in Cohen 2001, p56).

> "Programmes such as the proteome project are more diffuse, and without an obvious end-point, so one might question their usefulness, other than as a means of maintaining public awareness that the human genome sequence alone will not cure disease" (Blackstock and Mann 2001, pS1).

Despite the above misgivings, the Proteome Project was spurred on by the 'promise of proteomics' (Nature 1999). If the Human Proteome was mapped, it was believed that eventually it would help to identify new drug targets for specific diseases and individuals as a kind or personalised approach to medicine. This aspiration is reflected by Fields (2001): "there is much that genomics cannot do and so the future belongs to proteomics" (p1221). It would appear then that if scientists were not agreed on a large-scale proteomics project, they were agreed that understanding proteins is fundamental to understanding the human body. Indeed, Macbeath (2002) claims we can only grasp an understanding of complex organisms when we learn how proteins interact with one another.

Rather than any groundbreaking conceptual issues, the result was that the rise of proteomics stemmed out of necessity partly due to the limitations of genomics. Yet, throughout the shift in emphasis from genomics to proteomics, there has

always been an appreciation within the scientific community of the complexity of the task at hand, and also of the future problems that may lie in wait. In Chapter Five I theorise the *promise of proteomics* as a way of attracting funding and mobilising actors but also as a concept that if oversold may prevent stabilisation.

THE DEVELOPMENT OF THE HUMAN PROTEOME ORGANISATION

After the success of the Human Genome Project and powered by the advent of protein diagnostic technology during the 1980s and 1990s "large-scale protein studies seemed attainable" (Patterson and Aebersold 2003, p314). In February 2001, seven years after Wilkins' *proteome* articulation, and one week before the publication of the first Human Genome papers, the Human Proteome Organisation (HUPO) was established. HUPO was set up as an international consortium of national and regional proteomic research centres (HUPO 2005) and launched as a mirror image to its successful predecessor, the Human Genome Organisation (HUGO). The intention behind its conception was to follow HUGO and create a secure scientific infrastructure. HUPO and the HPP were created however under a cloud of uncertainty. The problem lay with the lack of support for the concept among factions of the community due to the complex nature of the proteome in comparison to the genome: "in terms of complexity, proteomics makes genomics look like child's play" (Service 2001, p2074). Thus, it was stated that: "HUPO will struggle to emulate its predecessor because human proteomics is not a single project with one endpoint that lends itself to HUGO style co-ordination" (Editorial 2001, p725). This acknowledgement of proteomics' multi-faceted nature adds credence to the view that proteomics is a networked based research activity requiring a myriad of experts in order to understand its complexity. In Chapter Eight, I analyse how this networked approach to proteomics impacts on scientific relationships and consider how interdisciplinary work is achieved.

The need for an international centralisation of proteomics work was paramount, since the cogency and genuineness of proteomics research was under scrutiny due in part to capacious dumping of disputable, unreliable and unverified protein data.

As Hanash (2004) states in his article, no single institution had the reserves, either financially or technologically, to deal with the Human Proteome single-handedly. As a consequence, in 2001, HUPO brought together international proteomics research centres under one centralised organisation. It was hoped that this alliance could assist in sifting out unreliable data and also "prevent companies locking up data under trade secrecy" (Kaiser 2002, p827).

On February 9th 2001, an international advisory council was unveiled bringing together experts in the field of proteomics from both the academic and industrial sectors (HUPO 2005). Over the next fifteen months, the council, in discussion with actors from industry, identified the major areas of concern in the proteomics field. Over a year later on the 29th April 2002, HUPO's advisory board declared that they had identified five key areas of human proteomics that HUPO would focus on. Hanash, who had been appointed inaugural president in June 2001, acknowledged that part of the reason for identifying these five specific incipient strategies was to find companies interested in funding the projects (as cited in Kaiser 2002). This statement is a further example of the importance of funding in scientific work[19].

HUPO's remit was as follows:

(i) the development of new lead technologies to quantify 5000 protein interactions,
(ii) the identification of abundant proteins in healthy adult human blood and the investigation of the influence of environmental variations, such as age and gender,
(iii) the systemisation of data and protocol standards for the heart and other existing proteome organ studies,
(iv) the development of a library of 50,000 high quality antibodies for every human protein, and
(v) the development of bioinformatics databases, analysis software and annotation standards for 2D gel electrophoresis, mass spectrometry and

[19] For example, the HPP has had substantially less funding than the HGP.

protein arrays either in inter-linked proteomic databases or in a large centralised database (Kaiser 2002; Merrick 2003).

For each field an expert was chosen and elected to chair the initiative. Rolf Apweiler from the University of Heidelberg and the European Bioinformatics Institute (EBI) was chosen to lead the bioinformatics field. It is this field which is the focus for the current study.

WHAT HAS HUPO AND PROTEOMICS ACHIEVED? THE STORY SO FAR

Although still in its infancy, HUPO has arguably proved some of its critics wrong by continuing to expand in size. An important ethos behind the establishment of HUPO was to develop the directives of HUGO, and to include more countries than those participating in the HGP. Six months after membership of the advisory board was unveiled, the first Annual HUPO congress took place in November 2002 at Versailles, France. Since then, congresses have been held in North America (Canada) in 2003, Asia (Japan) in 2004 and again in Europe (Germany) in 2005 (HUPO 2006).

HUPO has also been successful in terms of attracting funding and increasing membership. It has founded a centralised base at Montreal for its secretariat (opened in October 2005) and financial resources have come from organisations such as the National Institute of Health (NIH), Genome Quebec, Amersham Biosciences, McGill University and the Canadian Institute of Health Research (CIHR). Numbers attending conferences have also increased exponentially with over 2000 delegates attending the 4th World Congress in Montreal (HUPOST 2005), and Rolf Apweiler celebrating the success of the 5th Congress in Long Beach (HUPOST 2007). This rise in attendance has also mirrored the increased interest in the proteomics field at large, with Burrill claiming that from the early twenty-first century proteomics has evolved from a word no-one even discussed, to "the new darling of the investment community" (as cited in Service 2001, p2074).

This recent growth in the study of proteomics has been driven by the 'omic' and in particular the 'genomic' revolution of the 1990s. During the proceeding decade, proteomics evolved into an autonomous area of research with potential for future progress in diagnostics and health research. According to the Office of the European Union, the use of the term 'proteomics' has grown steadily. A search on the Internet from the period of 1st January 1996 to 15th July 1998 found 162 pages containing the word 'proteomics', whereas a second search from 16th July 1998 to 31st January 2000 found 2,799 pages containing the word (Office of the European Union 2005). In 2009, proteomics continues to grow, expanding to harvest proteomics specific journals such as, *'Proteomics'*, *'Molecular and Cellular Proteomics'* and *'Proteins and Proteomics'*, and helping to establish multi-national proteomics companies such as Wilkins' *'Proteome Systems'*. Nevertheless, in some quarters, proteomics is still not accepted as an autonomous discipline. McNally (2005), in her article using the Issue Crawler technique, portrays a weak web scale network between proteomics related web-sites. Significantly, she pinpoints the lack of links from mass spectrometer web-sites - a principal technology in the development of proteomics - to proteomics web-sites. This resistance to the development of proteomics could be said to be similar to some of the problems Educational Research has had in establishing itself in Germany (van den Daele and Weingart 1976). As Kuhn (1996) claims, a new theory (or in this case new research area) implies a change in the rules and customs governing normal science and as such can be met with some resistance and opposition.

Despite suggestions that proteomics has not been fully accepted, the newsletter of HUPO states proteomics is a fast growing field, and developments in infrastructure (its centralised base), technologies (the new quadrupole Tof – MS/MS) and training initiatives are aiding its stabilisation (HUPOST 2006). The developments in training initiatives are explored in Chapter Nine. Here different actors have different views. Significantly proteomics advocates believe that bioinformatics (one of the key areas identified by HUGO) is of central importance in its development (Blackstock and Mann 2001). I continue the primer by providing a brief historical account of the development of bioinformatics.

PART TWO - BIOINFORMATICS

INTRODUCTION

It is claimed that in 1987, the President of the bioinformatics company 'D'Trends', Dr. Hwa A. Lim, coined the term *bio/informatique*[20] (Rao 2004). Nearly twenty years later, the desire to translate raw (post-) genomic data into 'useful' knowledge[21] has meant bioinformatics has matured into a multifaceted discipline incorporating a host of scientific specialities, including molecular biology, proteomics, transcriptomics, genomics, computational biology and mathematical statistics (Fenstermacher 2004). This interdisciplinary area of research attempts to combine the digital codes of humans with the inner workings of machines. But what is bioinformatics, and how has it developed? Below are a number of marginally differing characterisations that have been used to categorise this area of research: Bioinformatics is:

> "...the collective term for data compilation, organisation, analysis and dissemination" (Lim; D'Trends),
>
> "...is the computer-assisted data management discipline that helps us gather, analyse and represent this information in order to educate ourselves to understand life's processes" (Persidis, 1999, p828),
>
> "...a discipline that 'derives knowledge for complex computer analysis of biological data" (Nilges and Linge; Institut of Pasteur).

Due to lack of clarity in the definition of bioinformatics the National Institutes of Health (NIH) Bioinformatics Definition Committee was set up in 2000 tasked to officially define the area. The committee characterised bioinformatics as the:

> "Research, development or application of computational tools and approaches for expanding the use of biological, medical, behavioural, or health data including those to acquire, store, organise, archive, analyse or visualise data" (Huerta *et al.* 2000, p1).

---

[20] The English translation of which is bioinformatics.

[21] In Chapter Seven I examine how bioinformaticians claim that they produce both knowledge and information.

The common consensus of all the above definitions is that bioinformatics is a discipline where biology and computer science merge to form a single research area that attempts to make sense of coded biological data. Yet what type of data does this refer to and how are the data actually generated?

During its evolution bioinformatics has been strongly associated with large online (post-) genomic warehouses. These databases and repositories are the interactive platforms used to represent gene and protein sequence information. Gene and protein sequence data generated from automated sequence technologies are deposited into public and private database systems, which act as a 'virtual bank' storing the data. Initiated by the onset of robotic protein sequence machines in the mid-1970s (Maxam and Gilbert 1977), the number of online bioinformatics databases exploded during the 1990s as a result of big biology projects such as the HGP. Since then, numerous data repositories have appeared around the globe, with some of the larger and more recognisable ones being GenBank (US) at the National Centre for Biotechnology Information, EMBL (Europe) at the European Bioinformatics Institute, DDBJ (Japan) at the National Institute of Genetics and SwissProt - a protein database based at the Swiss Institute of Bioinformatics (SIB).

The term bioinformatics however does not only relate to the tool that stores the data but also the science that that attempts to analyse and make sense of the data. The enormous strides in technological developments during the last twenty years have meant that mass amounts of data are generated and deposited in these banks. In 2002, more than 23,300,000,000 bases of DNA existed in the public domain with databases doubling in size every nine months (Miller and Attwood 2003). By 2005, the three main databases GenBank, DDBJ and EMBL contained over 100,000,000,000 DNA bases and stored completed genomes for over 165,000 organisms (Mehnert and Cravedi 2005). This omic explosion can be characterised by a 2005 EMBL-EBI Press release, which states that there are large, open-access database for virtually all types of biological information (EMBL-EBI 2005a). The result of this data deluge has meant that some of the most pivotal roles in post-genomic science are those of algorithm developers and mathematicians who design bioinformatics programmes that help scientists compare and verify these

deposited sequences. Aspects of the role that bioinformaticians play within omic science and the ways in which the research field is stabilising into a vital component of genomic science are developed upon in Chapters Seven and Eight.

THE BEGINNINGS OF BIOINFORMATICS: A SHORT HISTORY

Bioinformaticians can present themselves in many guises; the algorithm designer, the programme designer, the biologist and the annotator. Indeed while bioinformatics continues to mature numerous other subdivisions are also beginning to emerge (Fenstermacher 2004). For example, there are different genomic and post-genomic sequencing techniques, and different scientific organisations and databases. There are therefore several strategies in place to universally standardise areas of the research (Ravichandran and Sriram 2005). Faced with the possibility that data repositories will spring up autonomously and segregate research areas, scientific communities are collaborating to develop standards for a single fixed data representation (Miller and Atwood 2003). This process of omic standardisation and the role that bioinformaticians play within it is analysed conceptually in Chapters Six, Seven and Eight.

By providing an historical account of the developments within bioinformatics and their connections to proteomics, I outline here the important events that have led to community standardisation and communication becoming a central issue in proteomics and bioinformatics. I begin by tracing bioinformatics back to semiotics, artificial intelligence and cryptography, and conclude by introducing the Proteomics Standards Initiative (PSI).

Modern day computer science can be traced back to the calculating devices of the 1930s and 1940s and the code breakers of the Second World War (WWII). In 1936, algorithm designer and British mathematician Alan Turing (1912-54) published on primitive computers that combined the theoretical and physical worlds. This was the birth of the *Turing Machine*; a technology that has become

the foundation of all modern theoretical perspectives about computers[22]. Modern day bioinformatics is also indebted to Margaret Dayhoff; the so-called 'grandmother' of bioinformatics (Xiong 2006). She was the first to use computers in biology in her book, *'Atlas of Protein Sequence and Structure'* (Dayhoff 1969), which used computer-writing software to compare protein characteristics (Smith 1990). In fact Dayhoff used old-fashioned "punched-card business machines to calculate molecular energies of organic molecules" as far back as 1947 (Moody 2004, p11). Introducing proto-computers, the computers that had just successfully broken German military codes (Knight 1997), into biology became a trademark of Dayhoff's later career in proteins and computers, and possibly the first understanding that proteomics and bioinformatics could have a strong bind. Chapters Eight and Nine considers the relationship that they have today.

The most important liberating communication platform of the twentieth and twenty-first century was also born out of a dispute cloaked in a cloud of secrecy and paranoia. In response to the USSR launch of Sputnik I[23] in 1957, the Advanced Research Project Agency (ARPA) was formed in 1958. This originally small agency went on to play significant role in US computer science (Abbate 2001). In the late 1960s ARPAnet[24] began using the theory of packet switching, constructed to abolish distance limitations in local computer networks (Roberts and Wessler 1970), and in 1969, it successfully linked communication between four major US universities (LivingInternet 2007). In 1971 ARPA became known as the Defence Advanced Research Project Agency (DARPA) and under this new address developed electronic mail. E-mail became the network's most used service and changed the application of the computer forever more (Abbate 2001). If Margaret Dayhoff was coined the *'Grandmother'* of bioinformatics, the ARPAnet could be coined the *'Grandtechnology'* of the modern day Internet and a major step towards an informational science.

[22] Combining the physical and theoretical worlds, the Turing algorithm bombe', was later to decrypt the German WWII electro-mechanical cipher machine called *'Enigma'* (Mackenzie 1996).
[23] The first artificial satellite to orbit the earth.
[24] The ARPAnet was a computer networked system.

We have learnt that the emergence of bioinformatics as a science began in the mid-1970s as a result of major developments in automating proteins, developments in DNA sequencing, the increased accessibility of computers and Margaret Dayhoff's work (Persidis 1999). In 1973, Cerf and Kahn began linking the ARPAnet to other networks - a type of inter-network (Abbate 2001). At the same time, High Energy Physicists (HEP) were beginning to utilise developments in networking facilities to subscribe to pre-prints generated from a central database. This was the foundation of online bibliographic databases and a precursor to current online scientific data warehouses (Gunnarsdottir 2005). Online databases are discussed in Chapters Six and Eight.

In 1975, Professor Doug Brutlag, later to be the founder of the first bioinformatics company '*Intelligenetics*', began studying sequences in molecular biology (Moody 2004). Brutlag wished to assist remote scientists by designing a system (SEQ) allowing them access to the sequences that the community was generating. SEQ was written as part of the MOLGEN (Molecular Genetics) project to be distributed via the ARPAnetwork. MOLGEN became one of the first online molecular databases with bioinformatics tools (Moody 2004). However, the group's efforts to develop the technology were curtailed when the National Science Foundation (NSF) refused to fund them because the ARPAnet, later to be called the Internet, was primarily used in defence research, not scientific research. Once more this highlights science's entangled relationship with funding which is analysed in Chapter Five.

By the late-1970s and early-1980s several groups were working on molecular databases in the US and Europe. In April 1982, the publicly funded European Molecular Biology Laboratory (EMBL) was created in Heidelberg, Germany, to co-ordinate molecular biology research in Europe. Four other auxiliary nodes have since opened in Grenoble, Hamburg, and Monterotondo in addition to EMBL-EBI outstation in Hinxton. During the same period, the National Institute of Health (NIH) funded Bolt, Beranek and Newman (BBN) to produce a US DNA sequence database to be known as the Genetic Sequence Data Bank (GenBank) that today contains publicly available DNA sequences.

In 1986 NSF set up the NSFnet to link all US universities and introduced the Internet into the scientific and academic domain. The same year UniProtKB/Swiss-Prot protein knowledgebase[25] was established. Swiss-Prot would later evolve into a major public protein sequence database distributed by the European Molecular Biology Laboratory (EMBL). The following year also saw '*IntelliGenetics*' win the contract to maintain GenBank. Under the BBN stewardship, the database had experienced numerous problems in attempting to cope with data quantities. *IntelliGenetics* and EMBL promised to improve the efficiency of the database (Moody 2004). Today GenBank, EMBL and DDBJ make up the International Nucleotide Database Collection. In 2005, this global network had collected and distributed 100 gigabases of sequence data (Mehnert and Cravedi 2005).

Towards the end of the 1980s two further events occurred that strengthened the growth of biological information databases. In 1988 the European Molecular Biology Network (EMBnet) was established (EMBnet 2006) as a network linking European laboratories using bio-computing, bio-statistics and bio-informatics in molecular biology research. The idea was to link local (sometimes isolated) nodes to a centralised national facility (Attwood and Parry-Smith 1999). This illustrates the network community of bioinformatics that is explored in Chapter Eight. While in the US the National Centre for Biotechnology Information (NCBI) was created as a national resource for molecular biology information based at the National Institutes of Health (NIH). Today the NCBI produces cutting-edge research in computational biology, while promoting and developing standardised computer software tools for genomic and post-genomic data analysis (NCBI 2005).

Each of the described events signified that bioinformatics was becoming a global phenomenon and an essential tool in molecular sequence research in the twenty-first century.

[25] This was an annotated protein sequence database.

BIOINFORMATICS SOFTWARE PROGRAMMES

'Visionaries' such as David Lipman directed the NCBI's activities outwards. They created the Basic Local Alignment Search Tool (BLAST) computer programme as a heuristic resource and methodological tool built to aid researchers compare new individual molecular sequences with a centrally stored established sequence (Altschul *et al.* 1990). By calculating the statistical significance of the similarity of two molecules - a fully annotated sequence in the main database and the new undisclosed entry - the programme was able to verify sequence alignments.

The BLAST algorithm was based on three simple steps: scoring, searching and mathematical significance and remains one of the most important bioinformatics software programme for analysing data (Korf, Yandell and Bedell 2003)[26]. The server was located in the same organisation running GenBank so that every query could be run against the NCBI database.

In the late-1980s the Internet was still difficult to use because of limited browsing and graphic capabilities. This began to change in the early-1990s when Berners-Lee together with Cailliau from the European Centre for Nuclear Research (CERN) devised the WWW (Castells 2000). The WWW programme was created on a Nexus computer and developed some of the theories of the computer hacker's culture of the 1960s and 1970s. By creating the HTML (HyperText Mark-up Language) coding system, Berners-Lee produced a more manageable and user-friendly information resource (Berners-Lee 1989).

BLAST launched their programme on the WWW so it could be accessed by personal computer users. The greater ubiquity provided by the WWW meant BLAST was able to reach larger groups of remote scientists (Moody 2004). Access and user-friendliness was further intensified when the Mosaic graphical web browser (1993) improved the presentation of bioinformatics programmes. Since its launch, BLAST has grown to foster a number of subsidiary versions

[26] It could be claimed that the BLAST algorithm is the skeleton standard in the field. See Chapter One on the discussion of the QWERTY keyboard and Linnaean taxonomy as skeletal standards.

including 'blastp' - a programme for the sequencing of protein on protein data and used by many of the major global bioinformatics organisations. At the centre of this are many original ideas laid down by Dayhoff in her pioneering work on protein sequencing.

THE EUROPEAN BIOINFORMATICS INSTITUTE

The World Wide Web (WWW) revolutionised the networking ability of the science community. In 1993 the Expert Protein Analysis System (ExPASy) World Wide Web server was launched as the first WWW server in the field of life sciences (Gasteiger *et al.* 2003). It was the one hundred and fifty-first website to appear on the web[27], and today is home to over six million visitors. Ron Appel, the founder of ExPASy, expressed the importance of the WWW in the development of the server:

> "Putting this together we thought we should allow other people to access the data. I looked around at what kind of systems we could use to achieve this and in July 1993 I found out about the World Wide Web" (Ron Appel as cited in Brewis 2005, p15).

ExPASy is provided by the Swiss Institute of Bioinformatics (SIB) and today includes the databases SWISS-PROT, TrEMBL, PROSITE and SWISS-2DPAGE, all of which are dedicated to the access and analysis of protein and proteomic data (Gasteiger *et al.* 2003). The principal benefit of the server is that it links databases and servers. It has now evolved to foster similar sites in places such as Australia, Canada and the US. As Appel implied in the previous quote, this would not have been possible without the platform provided by the WWW. Bioinformaticians uses of the WWW are examined in Chapters Seven and Eight.

The WWW not only revolutionised the networking ability of the science community, but information technology was also accelerating the expansion of bioinformatics as a research activity. Fuelled by the expected data deluge from the Human Genome Project (HGP) and fulfilling one of HUGO's primary wishes, in the mid-1990s bioinformatics underwent a "period of explosive growth and

[27] This further illustrates the close links that science and technology had with the development of the Internet.

development" (Boguski 1994, p383). This was reflected in Europe when, in 1994, the European Bioinformatics Institute (EBI) was opened in Hinxton, UK.

The EBI is a non-profit academic organisation located at Hinxton Hall, and is home to one of the major HGP funding agencies, the Wellcome Trust Genome Campus. The campus is also home to the Sanger Centre, named after Fred Sanger who sequenced the first protein in 1955 in Cambridge, and the UK MRC Human Genome Mapping Project Resource Centre (Attwood and Parry-Smith 1999). The EBI is one of the four outposts of the EMBL (Emmert *et al.* 1994) and is the European equivalent of the NCBI. One of the major activities of the institute is the development of the EMBL Nucleotide Sequence database; Europe's major nucleotide sequence data system and a further collaboration between GenBank and the Database of Japan (DDBJ). The institute places sequencing and bioinformatics together by maintaining the SWISS-PROT protein sequence database, and providing free data and bioinformatics services to all parts of the scientific community (Moody 2004). Its mission statement includes supplying data to all the scientific community free of charge, providing advanced bioinformatics training to all level of scientists, and disseminating cutting-edge technology to industry and academia (EMBL-EBI 2005b).

The institute is home to over three hundred people from different technical and geographical backgrounds and is split into seventeen separate groups. These include the microarray group, the computational genomics group, the sequence database group and the newly-created proteomics service team. The sequence database group is an amalgamation of activities related to both nucleotide and protein sequence databases. The group is headed by Rolf Apweiler, chair of HUPO's PSI, and president-elect of (HUPO) from January 2007 and has five projects including the database Uniprot; a central database for protein sequence and function produced by combining information in Swiss-Prot and TrEMBL.

The proteomics service (PS) team is headed by Henning Hermjakob and provides the apparatus for the deposition, administration and analysis of protein and proteomics data. As part of its resource, it is home to the Protein Identification Database (PRIDE). This is an open source project that holds nearly 200,000

protein identifications and can be accessed or downloaded freely via the web. This database is arguably the first experimental 'proteomics' database and its aim is to produce accessible data (Martens *et al.* 2005). In addition to the Pride project, the PS team have also created a centralised query interface software programme called Ontology Lookup Service (OLS), which controls and standardises proteomics vocabulary. The regulation and standardisation of proteomics data is key for the proteomics service team. This is highlighted by the significant role the team plays in co-ordinating the European contribution to standardise proteomics data. The PRIDE database is closely coupled with the Human Proteome Organisation Proteomics Standards Initiative (HUPO-PSI) and it is hoped controlled vocabularies constructed by the PSI will increase the effectiveness of the protein database. Research conducted at HUPO-PSI informs the analysis in Chapters Five to Nine.

THE PROTEOMICS STANDARDS INITIATIVE

The Proteomics Standard Initiative (PSI) was established in April 2002 in Washington as part of a working group of the Human Proteome Organisation (HUPO-PSI). Its purpose is to "define and promote community standards for data representation in proteomics and to facilitate data comparison, exchange and verification" (Orchard, Hermjakob and Apweiler 2003, p1374). The PSI emphasise that the process is an open and inclusive involving actors from academia, industry and business. During its first meeting there was recognition of the continued fragmentation of deposited protein data into online warehouses and the need to standardise proteomics to overcome the problem. Data standardisation was seen as imperative for quality control, especially within an emerging discipline where the technologies generate a number of false-positive and false-negative results. In this sense the initiative was viewed as fundamental for the development of proteomics by vetting generated and deposited data (Orchard *et al.* 2003).

At the time of fieldwork, the PSI were active in three broad areas: the Molecular Interaction Standard including Protein-Protein Interactions (PSI-MI): the Mass Spectrometry Group (PSI-MS) and the General Proteomics Standard (GPS),

which includes the representation of overall proteomics experiments (Hermjakob *et al.* 2004). The ultimate aspiration of the PSI is to fulfil one of the major reasons for the formation of HUPO in the first place. The PSI was set up to prevent the lock-up of proteomics data inside local organisations. HUPO's aspiration is to standardise the whole proteomics field so data can be read, deposited, retrieved and analysed further by scientists operating in the discipline from anywhere in the world. This is the remit of the GPS group and their mission is:

(i) to construct a standard representation of proteomics data,
(ii) to standardise the discipline's ontology,
(iii) to negotiate an agreed minimum level of report detail (Taylor *et al.* 2006).

The standard format for detailing a proteomics experiment replicates the established microarray community and their Minimum Information About a Microarray Experiment (MIAME), found at the Microarray Gene Expression Database (MGED) group (MGED 2006). The hope is to create commensurate data reporting guidelines across post-genomics communities.

SUMMARY

This chapter has described the historical developments of proteomics and bioinformatics throughout the last century. It provides a timeline that records the level of stabilisation of both research areas. It indicates that the terms 'bioinformatics' and 'proteomics' are still perhaps ill-defined, and as they begin to develop there is evidence that they are being subdivided, thus producing more professional roles. Despite this, the chapter describes how some scientific groups believe in the 'promise of proteomics' as a post-genomic science. Subsequently, the chapter has presented science as having different factions, with some actors having different social interests to others (reductionist versus holistic scientists). Moreover, we begin to see how science and technology are dependent on large amounts of funding, and as the Human Protein Index (HPI) story suggests, this is imperative to the successful development of a new research field. To this end, despite the chapter serving as a scientific primer, it has also highlighted the role of

the 'social' in science and prepares the reader for the proceeding empirical chapters.

In the following chapters I focus on the emergence and stabilisation of proteomics and bioinformatics. I also consider the relationship between funding and research, and unpack the different professional roles found within bioinformatics. Firstly though, I move into the second of the literature review chapters (Chapter Three). I begin with a review of the history and philosophy of science (HPS) and the social studies of science (SSS) to illustrate some of the functions of scientific standardisation, particularly those in relation to scientific stabilisation.

CHAPTER THREE: STANDARDS, BOUNDARY CLASSIFICATIONS AND PARADIGM SHIFTS

INTRODUCTION

The derivation of the word *standard* can be traced back to the 'battle of the standard', an early military contest between the English and the invading Scottish army at Northallerton, Yorkshire in 1138 (McArthur 1999). To celebrate a particular army's identity a flag was raised. The flag was called *the standard* to illustrate its substantial standing. Later, the term standard was adapted and adopted to represent the King's Standard meaning *the best* or *the optimum* (McArthur 1999). In science, the term standard can take many forms (Busch 2008; Eriksson and Webster 2008; Fujimura 1992; Stephens, Atkinson and Glasner 2008a; Timmermans and Berg 2003, Tournay 2008). Here I provide a brief, focussed exploration of the relevant literature on standards in biomedical science that will be forward referenced to the proceeding case study material.

Focussing initially on the origins of scientific method and scientific classification, I illustrate some of the debates that were pervasive around the seventeenth and eighteenth centuries, disputes left largely untouched and underdeveloped until being reconceptualised with the development of science and technology studies in the 1960s and 1970s (Barnes 1974; Kuhn 1962; Latour and Woolgar 1979; Mulkay 1979; Price 1965). To this end I begin with Shapin and Schaffer. Overall, however, the chapter (i) presents a brief history of scientific method and (ii) illustrates three accounts of classifications and standards as exemplars depicting the position standardisation occupies in technological, scientific, political and commercial developments. The first of these accounts is the Linnaean taxonomic classification model, the second is drug classification in the UK and the third is the QWERTY keyboard design. In each example I demonstrate how the creation of standards and the construction of classifications

order particular social worlds by producing common platforms or measures that aid communication and comprehension. Finally, the chapter examines the significant role that standardisation performs in the authentication of autonomous, scientific disciplines by concentrating on the construction of boundary classification and Kuhn's notion of paradigm shifts. This is particularly pertinent in an era when standardisation is once more becoming a key component in the stabilisation of contemporary scientific research areas.

THE BEGINNINGS OF METHOD AS A STANDARD

The Royal Society of London established in 1660[28] led the movement from arcane and conceptual methodology towards more mechanical and empirical methods. Convinced by Francis Bacon's quest for new knowledge, and his belief that experimentation should be the light that would reveal all that was hidden in the universe, Charles II opened the Society. The Institute's interests were wide and varied, and ranged from attempting to understand the inner working of the body (biology) to comprehending the outer workings of the universe (astronomy). In light of this, the Institute became the home for the beginnings of 'small science'.

Conceivably, one of the most accounted-for and reported antiquated experiments of the Institute was performed by two of the Society's more distinguished patrons. In 1660, building on the invention of the barometer by Italian scientist Evangelista Torricelli, Robert Boyle and his assistant Robert Hooke invented their first Air-Pump machine (Boyle 1660). The experiment, which involved pumping atmospheric air in an endeavour to quantify air pressure, was replicated and repeatedly publicised around Europe as the correct and proper way to conduct a 'natural philosophy' experiment. Resultantly it was held up as the gold-standard approach to conducting a scientific experiment[29].

According to Shapin (1988), the tale of the Air-Pump has since become emblematic, demonstrating how nature can be controlled through technological

[28] Today the Royal Society of London is known as the Royal Society.
[29] A standard which Fuller (1997) believes does not exist in the social sciences: 'This leaves the overall impression that the low acceptance rates in social science journals is due just as much to a divergence in standards as to a surfeit of poorly crafted articles' (p21).

experimentation in order to generate scientific information. However he also demonstrates how the Air-Pump test was a rather isolated occurrence of an assay in action during the seventeenth century. He claims that the number of Boyle's contemporaries eager to view this unique contraption was small in comparison to the large numbers who could not attend the viewing (Porter 1995; Shapin 1988). Nevertheless, despite the comparatively low numbers who viewed the experiment, Porter (1995) insists that Boyle's great belief in empirical enlightenment, over and above the virtues of theory, meant that the number of advocates of experimentation began to increase. Accordingly, the growth in experimentation was not a result of first hand observation, since the obvious travel difficulties in the seventeenth century meant many of Boyle's peers could not view his demonstrations. Instead, its growth was a direct result of articulate reporting of the procedure of scientific method. Therefore, it is perhaps Boyle's verbosity, rather than his contribution to pneumatics that science can be most thankful for. His discursive manner often meant that he wrote in-depth detailed accounts of the experiments he was performing. His imagery was so convincing and his reporting of the event so thorough that actual 'live' witnesses were regarded by most as unnecessary (Shapin and Schaffer 1985). In Chapter Six of this thesis I examine the importance of documentation within post-genomic science. Focussing on the PSI proteomics data reporting standards, I argue that documented proteomics standards aid both scientific communication and disciplinary stabilisation, while also help to promote the professional identity of hybrid-like scientists.

Counter to the claim that Boyle's reporting had the support of many, Shapin and Schaffer (1985) maintain that Boyle's assertion of an irrefutable and incontrovertible method of collecting data was challenged by the materialistic philosopher, Thomas Hobbes. They state that Hobbes considered experimentation to be inherently private and that only a few people could view it at any one time. Subsequently, Hobbes did not consider experimentation as a universal phenomenon; a key principal of modernity for the enlightenment theorists. This opinion is echoed by Harry Collins (1988) who holds a similar reservation to Hobbes regarding experimentation. Collins states that: "where possible, experiments are still done in private because, the initiated aside, confidence in the facts will not survive a confrontation with Nature's

recalcitrance" (p727). Once more we see a further argument for the documentation of standards - as a way of disseminating so-called 'facts' to others. In the case of Boyle, Hobbes claimed that the Air-Pump actually leaked and that numerous pre-trials of the pump did not work leaving the results, in his mind, imperfect, flawed and shrouded behind the notion of experimentation (Shapin and Schaffer 1985).

Despite subsequent authors (Porter 1995; Shapin and Schaffer 1985) supporting the basis of Hobbes' critique, history suggests that Hobbes' personal attack on Boyle's experiment and his generic critique on the ethos behind experimentation failed dramatically. Boyle's influence was such that by the eighteenth century, "experimental knowledge" had to a sizeable degree "come to be defined in terms of potential reproducibility" (Porter 1995, p15). As Fuller (1997) states: "in the short term, Boyle won and Hobbes was excluded from the Royal Society" (p21).

Some writers and analysts have attempted to explain the Boyle/Hobbes outcome in terms of *social* relations. A closer look at the social standing of the two men reveals Boyle as a respected *gentleman* of English society (Porter 1995). Thus, there has been a school of thought (Barnes 1974; Porter 1995; Shapin and Schaffer 1985) that suggests that because Boyle was a revered pillar of the community his notion of scientific experimentation withheld Hobbes' challenge. Whether this is the case is open to debate, but the ideas of community consensus, intellectual respect and trustworthiness are virtues that have since become the bedrock of the social studies of science (SSS). In fact, Shapin (1988) has been fascinated by both the physical and social settings that surround experimental science and he places sharp emphasise on the significant roles they play in the process of scientific stabilisation. I return to issues of community consensus and the significant role the community plays in *socially* shaping a discipline in the empirical chapters and differentiate between *core researchers* and *peripheral researchers*. In relation to Boyle and Hobbes, Shapin and Schaffer (1985) show how the public were utilised as witnesses to legitimate the experimental method in the seventeenth century. Demonstrations were performed in the houses of gentlemen whose second hand reports were trusted because of their respected position in society. This account is a clear example of the roles that the 'social',

the 'cultural' and the 'local' play within scientific practice since Shapin and Schaffer describe how Hobbes questioned Boyle on whether anyone could be a public witness, knowing full well as a non-member of the Royal Society he would be excluded. This 'sociality' in science is also captured by Livingstone (2002) who argues that science is part of culture and not distinct from it:

> "...For science is supposed to stand free and unconstrained above the messiness of local circumstances. I want to suggest, however, that science is not above culture, it is part of culture. Science does not transcend our peculiarities, it discloses them. Science is not a disembodied entity; it is incarnated in human beings. Geologists do not shed their ethnicity when they engage in fieldwork; micro biologists do not discard their gender when they walk into a biotechnology lab, anthropologists do not set aside their politics when they pitch their tents among [a] forest people. Science is not some eternal essence slowly taking form in history, rather it is social practice earthed in concrete historical and geographical circumstances" (Livingstone 2002, p10).

Like Livingstone, this thesis supports the view of scientists as social actors. Much in the same way Shapin and Schaffer (1985) demonstrate that scientists were more like engineers than priests, ingenious craftspeople who manage and manipulate designed workspaces, this thesis demonstrates how big biology (Hevly 1992) – a significant shift from the small science of the Royal Society of London - is formed of collaborations between numerous researchers who manage and manipulate one another in order to solve shared biological problems. To this end, scientific endeavours such as the mapping of the human proteome have transcended national and cultural boundaries to become global partnerships. In particular, Chapters Eight and Nine demonstrate how researchers cross national and disciplinary divides to sort out these shared problems of research, while the focus of this thesis is on the social in science.

LAW AND ORDER: THE NECESSITY OF METHOD

While Shapin and Schaffer (1985) illustrate how scientific method evolved into the accepted way to perform science, others have directed their attention to shed light on the reasons for this procedure. In his book *'Changing Order: Replication and Induction in Scientific Practice'*, Collins (1992) presents the example of a

well-known comedy sketch from the TV show 'Monty Python's Flying Circus'[30]. The scene revolves around the mistranslation of phrases from Hungarian into English and back again, and is used to illustrate the basic principle that "without order there can be no society" (p5). What Collins (1992) aims to demonstrate is that without some sort of fundamental, shared and uniform understanding within a community, then the community will disintegrate into disorder. In the Monty Python sketch, what should have been a habitual and ordered interaction between two men spiralled out of control, eventually ending in a bar brawl. According to Collins, this was a direct result of a lack of common order to the conversational encounter.

The need for a sense of mutual order in society so that people can understand one another is clearly evident in science. Collins (1992) states that: "science like any other cultural activity rests on a foundation of taken-for-granted reality" (p18), which in turn lays the foundation for *inter-* and *intra-* disciplinary communication. Therefore, it could be argued that scientists study science through an ordered frame of reference which, beginning in the seventeenth century, has evolved through history to become the standard way to perform scientific practice. This common ground of understanding is what Merton (1943) calls the 'norm of universality'- a sense of ordered reproducibility which was purportedly initiated as a result of Boyle's Air-Pump experimentation. In principle, what Merton (1943) is suggesting is that science has an order whereby, in theory, anyone can use a *'guide of action'* in order to replicate the work of those that preceded them[31]. In turn, this can be used to attempt to reproduce results or validate claims. This source of replicability is determined by, and dependent upon, a standardised and uniformed action that has become known as experimental methodology. The notion of *inter-* and *intra-* disciplinary communication between post-genomic communities is examined in Chapters Six and Eight where I focus on the construction of (inter-) community-based standards.

[30] The Monty Python show was a British TV Show elevated to iconic status as part of popular culture.
[31] This would support Jordan and Lynch's (1998) work on recipe-type knowledge.

As Law (1994) further argues, research is a method of ordering, and scientific research is a particular activity that orders a means to discovering '*truth*'. Moreover, if science is the method to uncover '*truth*', then empirical methodology is the ordered technique that regulates the uncovering discipline (science). It is Boyle's (1660) original report of the Air-Pump experiment that has created an ordered method enabling science to *construct* a standard base to compare discoveries. This social infrastructure (scientific method) also provides science with a social foundation to be able to support a juncture of social organisation and order (Bowker and Star 2000). Whether it is the standardised assay or the homogenised report, the ordered procedure of performance acts as an institutional yardstick to measure one theory against another. This *standard* gives science a convention of communication, comparison and replication that is arguably unparalleled in other knowledge claims and ideological frameworks (for example religion or magic), and it is the work of standards that has been of particular interest to science scholars.

CLASSIFICATIONS AND THEIR CONSEQUENCES: THREE ACCOUNTS OF STANDARDISATION

The construction of boundaries and the creation of disciplinary standards are crucial in the development and '*trajectory*' of new specialisms (Bowker and Star 2000). Indeed, as an ordering technique, classificatory systems and standards are two sides of the same coin (Bowker and Star 2000). In this section I provide three separate accounts of standardisation and classificatory systems - Linnaean Taxonomy, the UK drug classification and the QWERTY keyboard. Together, they analytically anchor the thesis with the significant contributions presented in the empirical work of Chapters Five through Nine. The problematic relationships between 'natural' and 'social' categories, taxonomies and classificatory orders are discussed, and are directly linked to the proceeding empirical chapters. This provides the context for the later analysis on how contestations over disciplinary classifications and community standardisation endeavours are securing coherence for the emerging specialisms of proteomics and bioinformatics.

A century after Hobbes and Boyle were deliberating the purity of scientific order, Swedish physician, Carl Linnaeus (1707-1778), gave the world of flora and fauna a new taxonomy. Whereas the natural philosophers were still debating the merits of correct method, Linnaeus had published his tenth edition of the *'natural system'* (Linnaeus 1758); a classification system that attempted to order the natural world. His work had an enormous influence on the knowledge claims of nature with one contemporary colleague, in particular, stating that: "it has been said that God created nature and Linnaeus gave it order" (Schiebinger 2000, p12, Fara 2003, p19).

Linnaeus's religious beliefs followed the same teleological path as the naturphilosophen. As a deeply religious man, he believed the study of nature would reveal the divine natural order of God's creation (Krefting 2005). His taxonomy, however, was criticised by several members of the *scientific/natural world* community who believed he had chosen "an arbitrary plan rather than one that was divinely ordained" (Fara 2003, pp20-21). Despite this, Linnaeus believed his task was to construct an ordered classificatory system (Schiebinger 2000). In the 19th century the Linnaean binomial system of reporting emerged as the standardised system for classifying species and assisted in the stabilisation of taxonomy as a research field (Schiebinger 1996, 2000). Similar to the Linnaean example, in this thesis I show how classificatory systems and standards are social constructions created to order data and knowledge in particular post-genomic specialisms. Furthermore in Chapters Five, Six and Eight, I demonstrate how the construction of proteomics standards is aiding the identification and stabilisation of the proteomics research field.

Linnaeus intentionally used Latin to name the species because he considered it to be the universal language of comprehension, and with it invited the world to embrace his *universal* classification (Jenkins 1978). He also 'lumped' together species in order to reduce the number of categories. The idea of lumping and splitting is a well known issue for any discipline attempting to create rigorously defined categories. The difficulty occurs when you create categories and need to assign examples to them such as the classification of genetic disorders (McKusick

1969). Certain aspects of lumping and splitting are also discussed in Chapters Six and Eight in relation to the construction of ontological categories.

The end result was that Linnaeus produced a system of order that was universal in its utility and which was believed to be infinitely better than the preceding free-standing disorder. Linnaeus' classifications were artificial boundaries loosely based on observational techniques. However, his classifications did offer the possibility for a type of shared understanding of the natural world by providing order to a previously unordered scientific discipline. In Chapters Five and Six I analyse the construction of new boundaries and classifications, paying particular attention to the boundary construction of scientific disciplines.

The debate around the rationale of classification in taxonomy eventually divided the field. The twentieth century saw the emergence of experimental methods of taxonomy that were believed to be more '*scientific*' and more '*objective*' than their predecessors, challenging the orthodoxy of the Linnaean-based style (Heslop-Harrison 1953). According to some authors, the rigid and fixed Linnaean categorisation of species was unable to incorporate the new ideas of Darwinian gradual speciation (Dean 1979)[32]. Criticisms of the Linnaean model suggested that at the very least a component of flexibility was missing from his classification of species[33]. In Chapter Six I discuss and analyse how the creation of emerging standards necessitate that they are mutable, elastic objects (Bowker and Star 2000). For example within the PSI, *standard creators* write into the format opportunity for further changes to take into account technological developments or shifting foci (p171).

In the case of Linnaeus, Darwin even went as far as to argue that the ideas of a species classification were invented "fictions of fallacy in the taxonomist's mind rather than [any] objectively existing entities in nature" (Dean, 1979, p216). It appeared then that tension was surfacing between whether species classification

[32] In a similar way to how the field of taxonomy was partitioned, Chapter Seven of this thesis demonstrates how a division is emerging in the field of bioinformatics between *bioinformaticians* and *bioinformaticists*.

[33] See Hanseth, Monteiro and Hatling (1996) on the tension between standardisation and flexibility.

was a *naturally detected* or *socially constructed* phenomena or "whether classification is a process of invention rather than discovery" (Dean 1979, p212). The upshot of this challenge, according to Constance (1951), was that the more objective methodological approach of *experimental* taxonomy attempted to improve upon the low standing of taxonomy as a scientific discipline by rationalising its techniques. Therefore, not only do the classifications that Linnaean produced have pertinence to how ontological standards are constructed in proteomics, but the development of taxonomy itself shows symmetry to how I explore the ways in which proteomics and bioinformatics are beginning to stabilise into identifiable research fields.

Updated with the incorporation of biosystematic techniques, the Linnaean system remains the skeleton for twenty-first century biological classification. This longevity is perhaps the result of the need for a standardised and ordered classification that could aid community comprehension proving more important to the biological discipline than whether the nomenclature was based on any synthetic or intrinsic foundation. Likewise, in Chapter Six the rationale for the construction of proteomics standards is considered from the viewpoint of whether a standard should emerge via community consensus or whether there is a more urgent need for one to be created by whatever means (pp175-176).

Bowker and Star (2000) argue it is human nature to classify and standardise in order to provide benchmarks for further replication and order. Standards are the 'social basis' of scientific discovery because without communication or comprehension there is no community – the basis of scientific activity (Collins 1992). To this end, the Linnaean account demonstrates how Linnaeus' categories have provided a common platform which scientists have used to communicate with one another for over two-hundred years[34]. In Chapters Six and Eight I explicitly concentrate on 'omic' communication and interaction as a form of scientific stabilisation.

[34] Interestingly, it is also a specific technique reintroduced into twenty-first century science where bioinformatics is playing a pivotal role in shaping taxonomic biological futures (Hine 2006, Mackenzie 2005).

Similar debates over whether classifications are synthetic or natural configurations remain evident 250 years later. A good example is the classification of *illegal* recreational drugs in the UK. In January 2006, the then Home Secretary, Charles Clarke, proposed an overhaul of the 1971 Misuse of Drugs Act in the UK (Travis 2006). The categorisation of illegal substances into classes A, B or C[35] shifts depending on factors such as the harm they are perceived to have on the individual and on the corresponding institutions' ability to police their use[36]. These classes can be both constructed and contested at any given time. Therefore it is palpable there is no *'natural'* class A or *'natural'* class B, and their perceived harm changes across boundaries of scientific evidence, government policy and social acceptance. Equivalently, in Chapters Five and Six of the thesis I show how the classification of proteomics and protein research has changed over time and I give evidence of how different experts from various disciplinary boundaries are involved in the construction of proteomics standards.

Despite heterogeneous actors in the UK holding different views on perceived harm of recreational drugs, there is a level of social acceptance that a particular hierarchy of drug classification exists and that some drugs are perceived to be more dangerous than others. Furthermore, as recently illustrated in the media outpour and subsequent public condemnation of North Wales Chief Constable Richard Brunstrom after he claimed that: "ecstasy is a remarkably safe substance – it is far safer than aspirin" (Daily Mail 2008), it is currently social taboo to suggest the decriminalisation of all recreational drugs use. This brings to the surface the issue of individual 'mavericks' versus the social consensus when constructing and consolidating a standard. Chapters Five, Six and Eight shows evidence of how the construction of standards and disciplinary boundaries is a community-based activity negotiated between numerous individuals whose beliefs and ways of making sense of the world have to be brought in align with the rest of the community.

[35] With Class A drugs being perceived to be more harmful than Class C drugs.
[36] See Latour (1993) 'We have never been modern' for more on the entanglement of science and policy.

On the 25th March 2007, the Independent newspaper raised the issue of the recreational drug classification debate with a front page headline entitled 'The Great Cannabis Debate'. The newspaper ran an article about the United Nations' (UN) statement that there is a "growing threat to public health from potent new forms of Cannabis", and made the link between the consumption of Cannabis and adolescent Schizophrenia. This was in stark contrast to the paper's policy ten years earlier when it campaigned for the decriminalisation of cannabis, culminating in the UK government downgrading the legal status of cannabis from Class B to Class C (Owen 2007). The change of policy based on epidemiological data, allied to such words as 'growing' and the *Lancet* study which compiled an 'index of harm' for mood altering drugs, highlights the shifting and fluid nature of these classifications. In fact, the *Lancet* study "proposes that drugs should be classified by the amount of harm they do, rather than the sharp A, B and C divisions in the UK Misuse of Drugs Act" (University of Bristol 2007). Professor Blakemore, one of those involved in the study, admitted that "at present there is no rational, evidence-based method for assessing the harm of drugs. We have tried to develop such a method" (University of Bristol 2007). In response, Home Office Minister, Vernon Coaker, claimed that "we have no intention of reviewing the drug classification system" (BBC 2007).

It is clear to see from this short account of recreational drug use that there is a reluctance to change existing and established classifications that have been put in place to order particular areas of life. This is one reason why it is important to examine standardisation and classification as it happens. It is also apparent that standards and classifications are socially negotiated and socially vindicated ordering devices, often constructed within a particular community yet substantiated by a wider society. In May 2008 the Home secretary, Jacqui Smith, ignored expert opinions and upgraded Cannabis from a class C drug to a class B. Here we have the division of two standard groups: *the standard creators* and *the enforcement agency*. Together with considering and analysing the ways in which proteomics standards are constructed by a particular proteomics community and substantiated by a wider user base, Chapter Six shows evidence of these two groups (pp169-176).

The third and final account I detail is the development of the QWERTY keyboard. Towards the end of the nineteenth century, the Remington Arms Company (1874) patented the QWERTY keyboard. The keyboard was said to be designed by Scholes and Dunsmore, and the name QWERTY reflected the order of the six left-hand letters in the top row of the keyboard's particular design. In 1905, an international meeting was arranged attempting to establish a standardised design for keyboards, which the QWERTY design won. Nearly 60 years later, in 1964, QWERTY was officially unveiled as the international standard keyboard and has since been documented in many other standards (ISO 4169, 1979; BS 5959, 1980; ISO 9241, 1998/9). Despite the keyboard being ubiquitous throughout most of Europe (except in France) and America today, Noyes (1998) suggests that the acceptance of, and to some extent dependence on, QWERTY designed keyboards is not based on any scientific or efficient rationale that it is the gold-standard type keyboard.

Noyes (1998) presents four possible reasons for the original design of QWERTY. One reason is that the inventors intentionally made sure that all the letters used to type the word '*typewriter*' were located on the top line of the keyboard. This was because when they endeavoured to sell the device to potential customers they could demonstrate its capability by typing the word 'typewriter' rapidly. This tactic was a way of demonstrating to the customer the design's efficiency. However, according to David (1992), this tale demonstrates that the model was not based on any better or more efficient design for its users than any of its competitors. David (1992) asserts that: "by no means need the commercial victor in this kind of systems rivalry be more efficient than the available alternatives" (p139). Nevertheless, the QWERTY design has stabilised into the dominant technology in its field.

The development of QWERTY has analogies with this thesis since Chapters Five to Nine document and examine the stabilisation of proteomics and bioinformatics, placing sharp emphasis on the original construction of standards within the field. In addition, in the same way that QWERTY is a technological standard (Timmermans and Berg 2003), Chapter Six describes how constructed proteomics standards are beginning to be made compatible with proteomics technologies.

Since its inception, there have been over twenty formal challenges to the QWERTY keyboard, some of which have even been heavily critical of its fundamental design (Griffiths 1949). Despite the 1930s DSK design being believed to be the better system (Noyes 1998), the QWERTY design has continued to withstand these assaults and remains the dominant typing technology design[37]. Chapter Six explores how members of the PSI ago about creating a universal proteomics standard and how they attempt to withhold other competing proteomics standards (pp173-176). Issues of ineffective computing designs and platforms are also discussed throughout Chapters Six and Eight.

The QWERTY design has gone on to stabilise to such an extent that today it is difficult to displace. Following Bijker's (1995) terminology, a type of closure has occurred where a process of consensual agreement has transpired, possibly during the 1905 meeting and re-established during the 1964 international standardisation meeting, which is difficult to re-open. Even Gould (1991), an expert on Evolutionary Darwinism, questions the powerful acclaim of what he calls the 'mindless evolution' of QWERTY, simply because it won an international competition over one hundred years ago. Against this background, it is clear why it is important to examine standardisation as it is happening. The beauty of this thesis is that I am privy to the actual process of standardisation. The three accounts discussed in this section – *Linnaean taxonomy*, *UK drug Classification*, *QWERTY keyboard* - suggest that the construction of the initial standard is highly important, since in each example the original skeleton of the particular standard has remained dominant. In Chapters Five to Nine, I examine the early signs of standardisation and stabilisation in the research fields of proteomics and bioinformatics. The painstaking yet captivating work of examining standardisation techniques demonstrates how categories, classifications and standards spring from social, cultural and political contexts. Work in this area is also imperative if we believe that the original construction of a standard shapes the future of the field (Akrich 1992). This acknowledgement is the reason why STS writers have focused their attention on the day-to-day practices of scientific

[37] Bowker and Star (2000) demonstrate how the construction of standards often creates winners and losers.

research when highlighting the social construction of scientific practices, and coupled with the importance of the origin of the construction and the opening of the *boxing process*, these issues are built on specifically in Chapters Six and Eight.

STANDARDS AND CLASSIFICATIONS AS A PROCESS: THE BIRTH OF STS

Debates around the merits of the correct and rational way to execute a scientific method in order to discover 'truths' were rife amongst natural philosophers in the seventeenth and eighteenth centuries. In contrast, however, STS authors in the twentieth and twenty-first centuries have tended to focus on highlighting the *social* processes and *socio-political* debates that are involved in determining a perceived correct way to perform scientific method. This process involves a type of *authentic* standardisation in order to determine a ubiquitous '*correct way*' to perform a protocol, which may ultimately lead to the validation of a particular research area. As the three accounts of the Linnaean taxonomy model, UK drug classification and the QWERTY keyboard design have illustrated, this standardisation process is often negotiated and renegotiated through periods of conflict and consensus between numerous actors (Stephens, Atkinson and Glasner 2006, Stephens and Lewis 2008).

Many of the initial ideas about order and classification in the seventeenth and eighteenth century were re-established in debates about science with the advent of science and technology studies (STS). Latour and Woolgar (1979), argue standards and uniformity were fundamental for the behind-the-scenes functioning, and indeed, *social construction*[38] of science. They demonstrate how standards help scientists make sense of their observations and facts. These social agents (standards) have been further contextualised through analytical terms such as *standardised packages* (Fujimura 1988) and *boundary objects* (Star and Griesemer 1989), while the process of standardisation has been the concern of

[38] Notably, Latour and Woolgar (1979) removed the world 'social' in their 1986 edition to fit with their strong techno-constructionist beliefs.

many influential STS authors (Bijker 1995; Bingen and Busch 2006; Fujimura 1996; Law 1994; Tournay 2008).

According to Bowker and Star (2000), standards are closely affiliated with classifications because they can serve the same function; both are attempts to *order* a particular social world with the aim of aiding heterogeneous actors understand each other better. In Chapter Five I conceptually analyse the classification of the term 'proteomics'. But whereas classification involves the grouping of a category, Bowker and Star (2000) define a standard as "any set of agreed-upon rules for the production of (textual or material) objects" (p12). The process of agreement is also of particular concern in this thesis, and is dealt with specifically in Chapters Six, Eight and Nine. Furthermore, Bowker and Star add that standards have both a temporal and spatial reach, inasmuch as they often exist over time and space to form homologies. Within scientific practice, a standard may involve the standardisation of method or protocol (Jordan and Lynch 1998), the standardisation of technologies to produce the right tools for the job (Clarke and Fujimura 1992), the standardisation of languages to produce a particular vocabulary or ontology (Coenen *et al.* 2001), the standardisation of data (Abbott 1988) and the standardisation of regulatory practices (Eriksson and Webster 2008). These standards are often created within socially constructed boundaries and may be translated across them as '*boundary objects*' or when agglomerated may form a '*standardised package*' (Fujimura 1988); a way to perform things that helps to bring together and 'routinise' research in the discipline. While Chapter Five proposes the concept that proteomics is a proto-boundary object, Chapter Six demonstrates how the proteomics standard is also a boundary object and argues that genomics communities are better conceptualised using Fujimura's (1992) concept of 'standardised packages'. In this sense standards can order a research area by setting parameters that then serve to shape the discipline. In turn, this can aid the authentication and validation of the research discipline by giving it a source of objectivity and comparability.

The standardisation of a procedure or a package is also what O'Connell (1993) calls the 'circulation of particulars'; the dispersal of approved ways of doing things that becomes uniformly and universally applicable in different settings.

Thinking of standards in this way, it is clear how the construction of a *gold-standard* helps to bridge communication barriers between different actors situated in heterogeneous communities. To explore this matter fully requires further discussion on standards.

Law (1973) comments on the standardisation process of specialised scientific disciplines. He pays particular attention to the crystallographic and protein communities stating the "basic crystallographic methods were uniform across the community" (p285). This uniformity of method is an issue that is developed in Chapter Five when charting the move from protein chemistry to proteomics. Yet the development of standardised methods for protein crystallography depended on the adaptation of methods from both the crystallographic and protein communities: "methods were widely shared and the object of great interest throughout the community" (p285). His study was an example of *inter*-community standardisation to create *intra*-community comprehension. Further examples of these types of *inter* and *intra* protein community standardisations are illustrated in Chapter Six of this thesis (pp157-160). Law questions why some specialised disciplines achieve standardised disclosure through their shared methods, while others achieve it through shared theories[39]. Although he does not really answer this peculiarity, in his empirical case he illustrates that no matter how disclosure is agreed, the standards were adapted and constructed through a process of scientific solidarity and social consensus, in his case by using Durkheimian ideas of mechanical and organic solidarity. In comparison, in the case of proteomics I highlight the relationship between what I term *core* proteomics researchers and *peripheral* proteomics researchers.

Law's (1973) early study is built upon by O'Connell's (1993) and his summary of the standardisation of electrical units in the late nineteenth century. O'Connell (1993) states that electrical units standards were "forged through intense social interaction – by an international group of electricians, research physicists and industrialists who hotly debated rival standards for a quarter of a century before reaching settlement" (pp136-137). Moreover, what O'Connell's paper elucidates

[39] This idea has been developed further by STS writers with work on boundary objects (Star and Griesemer 1989).

is that in much the same way that standardised method for protein crystallography was constructed through interdisciplinary social consensus, the standardised unit of electrical measurement was agreed upon through consensus arising out of initial interdisciplinary conflict. In Chapter Eight I examine issues of both interdisciplinary conflict and collaboration. But what these examples illustrate is the processes by which standards emerge; through negotiation and renegotiation by members of a community or set of communities. I now move on to a discussion of communities.

In Chapter One I described how Collins (1992, 1999) believes the role of a community in science is fundamental and demonstrated how his arguement that if the scientific network is strong it will regulate the discipline by sifting out deviance. In turn it is through the process of standardisation that the community can align and normalise the network to follow a certain procedure (Hanseth, Monteiro and Hatling 1996). Community consensus and professional peer pressure can co-exist alongside the standardisation of method. Porter (1995) gives the example of Vannevar Bush (1890-1974) to illustrate the importance of community consensus within science. Bush was an American visionary and staunch science defender who played an integral mediating role between the US government and the scientific establishment around the time of World War II. Bush sought to keep science sacred from the ever-extending arms of US government and wanted to make it self-regulating by producing a boundary between scientific practice and non-scientific practice (Gieryn 1999). As Porter (1995) describes, Bush believed that in the unlikely event that scientific method failed to spot errors in scientific practice then an alternative way to validate work was through scientific community pressure, for example peer review. The scientific community then acts as both a perimeter and a percolator, reinforcing its boundaries while sifting out the bad science and scientists. This kind of science kinship within a discipline is also evident in Traweek's (1988) work on cutting-edge High Energy Physicists (HEP), who shows how scientific agreement is shaped through tacit interactions between scientists - another example of peer orientated work. Consequently, it may be argued that the first wall of defence in protecting scientific activity is the *method* (and standard), but if chinks begin to emerge in this outer protection then the institution of science is supported by a

further wall, *the scientific community*. What these 'supports' achieve is to make scientific practices, scientific theories and scientific endeavours *seem* more certain. Yet the paradox is that for science to progress it may have to violate its own methodological rules (Feyerabend 1993) and/or cross over the supported community divides. This moves the scientific practice into more uncertain waters; precisely the areas which are the focus of this thesis, particularly in terms of how scientists overcome the uncertainties involved in everyday practice.

To summarise, the literature dealing with standards and classifications has illustrated the ways in which standards are assembled, and how classifications are shaped and developed by a particular community often during periods of 'change' or 'crisis'. This is what this thesis terms the *identification stage* (p161). As Misa (1992) comments: "...in fact, a distinctive characteristic of scientists and technologists is their ability to resolve controversies and engineer consensus" (p109), and the primary technique used to do this is to *standardise*. In this regard this thesis also addresses how consensus is *engineered* through standardisation and communication (in particular see Chapters Six and Eight). Therefore the development of communities and the construction of standards go hand-in-hand.

I now move on to discussing Kuhn's (1962) theory of paradigm shifts, relating the theory to the possible paradigmatic shift from reductionist biology to omic biology since the mapping of the Human Genome.

SCIENCE SEGREGATION AND BOUNDARY CLASSIFICATION

Kuhn (1962) endeavoured to differentiate between the day-to-day practices of scientists' work and macro-scientific triumphs. In an attempt to illustrate how science functions, Kuhn (1962, 1996) argued that the beliefs and practices of scientists throughout history have always been directed by an over-riding, ubiquitous ideological paradigm. This paradigm is accepted by the scientific community as the dominant scientific accomplishment until another competing paradigm emerges to replace it during a *scientific revolution*. One of the unique aspects of Kuhn's theory compared to earlier Mertonian doctrines of science is

Kuhn's idea that science does not *progress* in a cumulative fashion. He argued that when one paradigm rises to displace the previous dominant paradigm of the time (a paradigm shift), not all the achievements of the previous paradigm are preserved. This was in stark contrast to earlier beliefs on the workings of science which had celebrated the cumulative nature of science whose progress was guaranteed by scientific *method.* For Kuhn, any scientific revolution in science involved an amendment of previous scientific practices and beliefs; even the method was not sacrosanct (Kuhn 1962, 1970). Interestingly, the notion of a scientific revolution and the gaps that exist between revolutionary stages gave scope for social constructionists to illustrate how science is not always constrained by strong scientific guidelines, but can be directed by socio-political factors (Knorr-Cetina 1981; Pickering 1984).

A contemporary example of this paradigm shift in biology might be illustrated by the movement from a *'reductionist approach in biology'* (Kellenberger 2004) to a *'systems biology approach'* (Hood 2003) or the so-called omic revolution. This is described in Chapters Two and Five. Fuelled by improvements in technologies and scientific techniques (Liebler 2002), the Human Genome Project (HGP) has been a catalyst in promoting a change in how biology is conceptualised. Keller (2000) writes that the HGP has made it impossible to ignore the rift between the reductionist gene-led approach and this new holistic view. In this sense the movement from viewing or understanding biology through individual and autonomous entities such as genes and proteins towards looking at networks and systems such as genomes and proteomes, could be accepted to be a Kuhnian-like paradigm shift and the word omic symbolising a "redefinition of how we think about biology" (Liebler 2002, p3) in the twenty-first century.

According to Kuhn's (1962, 1996) theory however, the shift between paradigms is not abrupt. Once a paradigm has been accepted as the dominant realisation, he argues that there is a period of stability where the paradigm then acts as a model for further successful research. This period is what Kuhn coined *'normal science'*; a time when scientists continue to work under a predominant paradigm by supplying further foundations for its practice (Kuhn 1996). Included in this day-to-day 'routinisation' of science may be periods of standardisation,

automation, commercialisation, organisation and infrastructure construction (Jordan and Lynch 1998; Keating; Limoges and Cambrosio 1999). Using this Kuhnian terminology, it may be argued that one instance of *normal science* activity within twenty-first century biology is the current sorting out and organisation of *omic* science (see Chapters Five to Nine). For example, stage one (genomics) has undergone further development to stage two (post-genomics), with the creation of sub-disciplinary omic communities such as genomics, transcriptomics and proteomics, all marked by institutional boundaries (Chapter Six). According to some authors this potential three stage development of omic science (with systems biology being stage three) is all part of the same expansion of the genomics view of science (Collins *et al.* 2003). This idea is consistent with Kuhn's second unique concept - the concept of incommensurability. Kuhn (1996) claims that science guided by a particular paradigm will be incommensurable with science developed under an alternative one, by which is meant that there is no common measure of the different scientific perspectives. This particular point is an essential component behind the infrastructure of omic biology, since if the creation of sub-disciplines such as transcriptomics and proteomics are to become essential communities in its further development, then the research fields will require a common measure of comparison in order to share methodologies, standards and puzzle solutions. According to Kuhn's theory (1970, 1996) this is achieved because they *are* part of the same paradigm and therefore part of the same way scientists understand the world.

Kuhn's incommensurability theory has been criticised by, among others, Doppelt (1978) who attempted to organise what he believed to be Kuhn's rather cluttered concept. Doppelt delineates Kuhn's incommensurability concept into four types; ideas and languages, the mode of observation and perception, the list of *puzzles* to be solved and the criteria of adequacy for scientific explanation (Stephens 2005). These are important distinctions because the situation of omic science is further complicated when twenty-first century biology is viewed as 'visions' rather than 'stages'. For example, McNally and Glasner (2007) illustrate the different perspectives and different debates surrounding the gene and the genome and the apparent movement into a 'post-genomic' era. While accepting that this movement could be called the 'extended genome' (Dawkins 1982), McNally and

Glasner (2007) argue that the original concept of the gene would have to be redefined, and compare what this thesis suggests are stages within omic science as contested visions based on the flexible ways scientists name and classify things (Fuller 1997). As a consequence, perhaps it is fair to suggest that the modes of observation and perception within different post-genomic communities are comparable, but that the concepts and languages alter slightly. This observation would be consistent with Toulmin (1970) who argued that re*visionary* changes in science occur far more frequently than Kuhn's more grand re*volutionary* idea (for example the apparent move from gene to protein), and that scientific change is dependent on a mixture of freedom, chance and the relationship between innovative individuals and the wider community. If Toulmin's interpretation is to be accepted, Chapters Six, Eight and Nine illustrate how inter-community standardisation and boundary permeation are key techniques used by scientists to order the infrastructure of omic science, and to amalgamate these separate visions into one clear and coherent perspective.

To summarise, in the same way that experimental method was deliberated upon and constructed to produce a common understanding within the scientific community in the seventeenth and eighteenth century, so specialist techniques and methodologies within sub-disciplines must be able to cross institutionalised boundaries (Star and Griesemer 1989). Objects and methods need to be able to fracture the newly created institutional boundaries in order to promote shared understanding. This thesis will demonstrate that in a period when interdisciplinary research is the *gold-standard* archetype way to perform academic science, then the organisation of a solid social network is fundamental to the construction and stabilisation of a new research field that contains heterogeneous actors (Chapters Five, Six and Eight). As part of this, the thesis also reveals how *boundary objects* such as community standards and *boundary personnel* cross traditional borders to forge new networks and new modes of collaborations. Using the notion of paradigm shifts, the final section of this chapter describes a short account of standardisation in proteomics.

A SHORT ACCOUNT OF STANDARDISATION IN PROTEOMICS

The laws of genetics owe their origin to Mendel, who expounded the fundamental principles at Brno in 1865 (Waller 2002). Once established, Mendelian genetics became an example of a Kuhnian 'scientific paradigm' (Kuhn 1962, 1970, 96). According to science realists, as the dominant theoretical framework of the time, Mendelian genetics was able to transform biology from an imprecise practice into an 'exact' science (Kellenberger 2004).

Interestingly, Mendel's work on genetics elaborated upon and advanced the work of Linnaeus (Lemaine *et al.* 1976). Waller (2002) tells how Linnaeus's task was "to pigeon-hole living forms into neat categories" (p140) – a technique which Bowker and Star (2000) might call *torque*. As Linnaeus grappled to make sense of the multitude of species that he was classifying, he became convinced that God could not have created them all at once. Linnaeus was certain that life evolved, but not in the way Darwin (1859) was to view evolution. Instead, Linnaeus came to believe that species must crossbreed to create new hybrids (Waller 2002), and it was this evolutionary image that Mendel developed upon in his experiments of 'plant hybridisation' and his calculations of inheritance.

Nearly three-quarters of a century after Mendel's discoveries, Kellenberger (2004) describes how physicists were being attracted to the study of biology. The American Scientist Emory Ellis and his German colleague Max Delbruck, whose early work was on Quantum Physics, began to study the fundamental role of inheritance within simple organisms. They jointly constructed and established the *standard method* for the field (Ellis and Delbruck 1939), while their pioneering work broke existing traditional scientific boundaries by daring to cross them. The generation of a *standard* by Ellis and Delbruck helped to render information comparable and consequently comprehensible for consumption. A further seventy-five years later, genes are today beginning to be viewed by many in relation to genomes, and biology has become informational (Chapter Two). Nevertheless, a similar trend is occurring whereby contemporary pioneers are venturing across established boundaries. In Chapters Seven, Eight and Nine I

examine the ways in which experts cross traditional disciplinary boundaries to create new standards in proteomics and bioinformatics, paying particular attention to how mathematicians and computer scientists who have been attracted to the field of bioinformatics are able to collaborate with and comprehend biologists even though they have different backgrounds and training (Chapters Seven and Eight).

According to Keller (2000) then, the twentieth century in science was the century of the gene, beginning with the rediscovery of Mendel's work on heredity and ending with a partially completed draft of the Human Genome. The European Molecular Biology Laboratory (EMBL) bank launched in 1982 contains a database of completed and partially completed genomes housed at the EBI. The EBI has attempted to improve the resource and keep it up-to-date by developing the Genome Resource Review. Most importantly, the review is a comprehensive and standardised resource for completely sequenced prokaryotic genomes (Kersey *et al.* 2005). Since the first deposition of prokaryotic genome data in 1995, information, knowledge and techniques have developed at a rapid pace (Brooksbank, Cameron and Thornton 2005). Thus, the Genome Review has proved to be an essential resource adding up-to-date detail and annotation, while the EBI and its associated authors have attempted to standardise the production of this type of data.

As is described in Chapter Two, sub-disciplines are still evolving and stabilising within this systems biology approach. According to Liotta and Petricoin (2000), proteomics "is being proclaimed as the next step after genomics" (p13) and, as with Ellis's and Delbruck's (1939) work on genetics and the EBI's Genome Reviews with Genomic data publication (2005), a major goal within the day-to-day practice of normal science is the standardisation of disciplinary work to aid communication, comprehension and collaboration. Therefore it is appropriate at this point to look at the Proteomics Standards Initiative (PSI), which is one of the groups taking a leading role in this endeavour.

With the establishment of the PSI, progress is being made to develop common standards for data exchange within proteomics (Orchard *et al.* 2003, 2005a).

Initiated in 2002, one of its more daring objectives has been to create a General Proteomics Standard (GPS) (Chapter Two). The GPS is primed with the responsibility of attempting to construct a robust, future-proofed standard representation for both data representation and method in proteomics experiments (Brewis 2006). The overall goal is to produce a clear, coherent and consistent way of reporting data, and a basic format of how to (re)present all the collated data. The structure of these documents will help to identify, verify, legitimate and regulate the research area. Researchers at Manchester University (Taylor *et al.* 2003) have produced one of the landmark papers in this area. Their PEDRo XML programme attempts to design a methodical and systematic approach to modelling and distributing proteomics experimental data. Part of the model also includes guidelines on how to report fully a proteomics experiment (MIAPE); a general standard that has followed the blueprint of its predecessor in the Microarray community (MIAME). These are what Tournay (2008) might describe as 'operators of standardisation' – reference points.

The MIAME document stresses the notion that it is a set of guidelines and not a formal specification. As MIAPE is based on the structure of MIAME it is also framed as a set of guidelines, with the PSI determined to integrate them into proteomics journals (Orchard *et al.* 2005a, 2005b). Additionally, as is the case with MIAME, it will also include recommendations for (i) controlled vocabularies, ontologies and nomenclatures, (ii) clear descriptions of the design and type of technology involved, (iii) measurements of the separate levels of data processing, and (iv) a lengthy experimental description set to a certain structure so that results can be made comparable (Brazma *et al.* 2001 for MIAME). It is anticipated that this type of social ordering will help proteomics stabilise into a recognisable and authentic research area.

CONCLUSION

It is apparent that many of the debates between seventeenth and eighteenth century natural philosophers about constructing a functional and rational methodology to 'uncover truths' are being re-born with the birth of new scientific disciplines in omic twenty-first century biology. Although they are auxiliary

ancestors of the dominant rational scientific method, the social construction of a *correct* and *dominant* methodology, and the reporting of that protocol are attempts to regulate sub omic communities and disciplines. The *order* that this regulation creates not only validates and vets any data produced (Chapter Six), but also facilitates communication between researchers and communities (Chapters Six and Eight). This necessity is intensified further with the introduction of bioinformatics which, without some sense of social ordering and foundation for communication between the biologist and the bioinformatician, may lead to knowledge and information becoming blackboxed in one discipline or profession (Chapters Seven and Nine).

Within proteomics, Orchard *et al.* (2003) states that standardised data exchange is perceived as being "essential for data comparison, benchmarking and quality control" (p16) of all aspects that aid the stabilisation of the discipline. If this is true, it poses many further questions concerning the processes involved when socially constructing the frameworks and boundaries around science, and why these frameworks need to be constructed in the first place. Without challenge and scrutiny, the identity and value of constructed standards remain invisible. As such, the PSI is making a concerted effort to advertise the inclusiveness of open community participation within the standardisation process (Orchard, Hermjakob and Apweiler 2004), and consistent with past literature (Eriksson and Webster 2008), it appears that consensus within (post-genomic) scientific communities is formed through the creation of standard platforms. However, further questions may also need to be asked such as:

(i) How are the standards diffused to local users?
(ii) Are standards integrated into the new technologies?
(iii) How flexible are the standards and are they able to adapt to the evolution of scientific knowledge?
(iv) Is the wider network of actors actually interested in getting involved in the process?

While more global questions may include:

(v) What other communicative techniques are used in interdisciplinary big biology to aid communication?

(vi) What other devices aid the stabilisation of a research area?

Although Chapter Six concentrates specifically on the first four of those questions, the overall thesis addresses questions five and six. Explicit in the account of all empirical chapters is the idea that standards act as media for improving communication and organising activities, and as such, are an essential stabilising technique. In Chapter Five the focus is on the classification of proteomics, and its development is described through the concept of the proto-boundary object. In Chapter Seven I explore the development of bioinformatics and in particular focus on knowledge transfer and communication between biologists and bioinformaticians. Chapter Eight continues to highlight interdisciplinary communication between omic researchers, while in Chapter Nine stabilisation is examined in relation to the creation of academic qualifications and disciplines. Thus, if during the literature reviews I have concentrated on notions of chronology (historical developments); in the remainder of the thesis I am interested in social geography (boundary construction). Additionally, if I began by commenting on some major disagreements in science (Boyle and Hobbes), the rest of the thesis demonstrates how scientific actors try and overcome differences and uncertainties in scientific practice to forge agreements (it shows how they achieve and understanding of the world). But before introducing these empirical chapters, I first describe the methodological process used to collect the data that informs them.

CHAPTER FOUR: METHODOLOGICAL REFLECTIONS: A SOCIAL SCIENTIST IN A NATURAL SCIENTISTS' SETTING

INTRODUCTION

The aim of this chapter is to describe the methodological approach used in this research project. The overall ethos, to borrow words from Hammersley (2003), is for the author to reflect on:

> "why they did what they did and its consequences, both methodological and ethical [and to]...make explicit for their readers how their research was done, and their own role in producing the findings" (pp344-345).

The chapter is split into separate sections that discuss what I consider to be some of the more interesting methodological issues encountered in conducting qualitative research in the areas of proteomics and bioinformatics. It begins with a brief explanation of qualitative inquiry as a comparative technique to the ideas of experimental methodology discussed in Chapter Three. This is provided in order to introduce the approach taken in this particular study, and to reflect on the processes of access negotiation, data collection, data analysis and reflexivity. The chapter concludes by explaining the use of the particular methodological technique chosen and discussing the researcher's role as a quasi scientific insider. Overall, the chapter is divided into the following sections: *Introduction; Types of Methodology*; *Negotiation of Access*; *Sampling; Ethical Issues*; *Site Visit*; *Semi-Structured Interviews*; *RSSDP course on Protein Bioinformatics*; *Elite Interviewing*; *Action Research*; *Email; Interactional Expertise*; *Analysis* and *Conclusion.*

Some of the debates about the virtues of experimental method that were common in the seventeenth and eighteenth century are discussed in Chapter Three. These

Principles of Modernity, including empiricism, observation, logic and evidence are still dominant in the twenty-first century, and are often used to describe the new aspirations within omic biology. On August 13th 2007, Channel Four (UK) televised a programme called 'The Enemies of Reason'. Hosted by Richard Dawkins, the programme attempted to critique so-called irrational pseudo-sciences such as astrology and spiritualism. Despite his claim that a number of *Enlightenment* principles are under attack by a new wave of, what he calls, superstitious, dogmatic beliefs (his assault was on alternative medicines), Dawkins' overwhelming message is that science has given the world tangible benefits, and that it is the experimental technique, which has been the tool that has enabled these developments to flourish (Chapter Three). In this fourth chapter, I reflect on the type of methodology and the methods used in this study.

QUALITATIVE RESEARCH AND SOCIAL CONSTRUCTIONISM

As illustrated in Chapter One, social actors, and for that matter ideological frameworks, are complex entities and processes that need to be studied using different methodological approaches from the experimental technique used in scientific experiments. This approach is evidently more qualitative than its quantitative counterpart. In comparison to the more experimental methodology, Qualitative Inquiry (QI) has a shorter history. Its importance was established in the work of the 'Chicago School', which in the 1920s and 1930s combined a *positivist* urban sociology perspective with small scale *case study* interactions (Denzin and Lincoln 2003). The local and in-depth studies produced by Mead, Park, Sutherland and others paved the way for a tradition of more qualitative research. In this respect, QI is a contemporary methodological technique. Nonetheless, as others have argued, it is still a group of methods with a strong ethnographic grounding (Atkinson and Hammersley 1994).

Since the 1920 and 1930s, the use of qualitative methods has increased across the social sciences[40]. As the methodology has gathered momentum so the varieties of

[40] Despite this, there has been a division between more quantitative methodological approaches in psychology and economics with the more qualitative approaches of sociology and anthropology.

techniques and method types have also increased. Developments in technologies such as computer programmes, audio recorders and photographic instruments have meant new techniques have emerged and old techniques adapted (for example web crawling, multi-media analysis and digital recording). This has further added to the diverse types of method that Van Maanen (1987) believes the term QI encompasses:

> "The label qualitative methods has no precise meaning in any of the social sciences. It is at best an umbrella term covering an array of interpretive techniques, which seek to describe, decode, translate and otherwise come to terms with the meaning, not the frequencies, of certain more or less naturally occurring phenomena in the social world" (p9).

But despite the interpretative flexibility surrounding the term, Atkinson and Hammersley (1994) state that, in practice, qualitative research will have at least some of the following features:

> - "a strong emphasis on exploring the ***nature*** of particular social phenomena rather than setting out to test hypotheses about them
> - a tendency to work primarily with unstructured data, that is, data that have ***not been coded*** at the point of data collection in terms of a closed set of analytic categories
> - investigation of a small number of cases, perhaps just one case in detail
> - analysis of data that involves explicit interpretation of the meanings and functions of human actions, the product of which mainly takes the form of *verbal* ***descriptions*** and ***explanations***, with quantification and statistical analysis playing a subordinate role at most" (p248).

In this thesis I explore the ***nature*** of the proteomics and bioinformatics worlds by using qualitative methodology. I use interpretive constructionist sociology that begins with the actor's perspective (data that have ***not been coded*** but which are supportive of the arguments made) and on completion of data collection, I then re-interpret the accounts into the analyst's perspective (see Collins 2008 for a discussion of this switch). By exploring social phenomena in these two social worlds I then categorise (using thematic codes and utilising the qualitative software package NVivo) how omic science is socially constructed. Although Lynch (2001) claims that: "social construction is remarkably protean" (p242); readily assuming different meanings and truths, I use the term social

constructionism to analyse how 'things', 'categories' and 'truths' are socially negotiated, socially validated and socially substantiated. Furthermore, Mason (2006) asserts that: "the particular strengths of qualitative research lie in the knowledge it provides of the dynamics of social processes, change and social context, and in its ability to answer how and why questions in these domains" (p16). Therefore it is these social processes and social changes within science which are my main foci as a social constructionist of science.

TYPES OF METHODS

Over the last thirty-five years there has been a plethora of social studies of science and technology projects. Gold-standard projects include Collins (1975, 2004a) and Traweek (1988) interviewing high energy physicists, Gilbert and Mulkay (1984) interviewing biochemists, and Suchman (1987) interviewing computer scientists. The discipline has also seen Fujimura (1987), Knorr-Cetina (1981, 1999), Latour and Woolgar (1979) and Pickering (1984) pioneer the ethnography of laboratory observation. In more recent times, other innovative methods have been used to explore the 'social construction' of scientific practice. These have included case studies (Rabinow 1996), imitation games (Collins *et al.* 2006) and the use of the *IssueCrawler* to map electronic networks (McNally 2005).

In this study, three main types of qualitative methods were utilised, mirroring the three separate stages of the fieldwork. They were (i) a site visit to a cutting-edge research centre, which I consider as a type of micro case-study[41], (ii) semi-structured interviews with scientists and scientific researchers working in particular universities within the UK, and (iii) participation on a RSSDP *Protein Bioinformatics* course, where I used many of the techniques used within ethnographic observations of laboratories. In addition to these methods I also:

(i) attended scientific workshops, scientific presentations, proteomics symposia and interdisciplinary conferences, which have been a source of further scientific information and also a vital part of the methodological process,

[41] Focussing in-depth on a particular group and their attempt to create a proteomics standard.

(ii) conducted a thorough literature review,

(iii) performed documentary analysis of the literature,

(iv) corresponded in several e-mail interactions, which included three extra participants answering a set of questions relating to standardisation in scientific practice and the concept of 'systems biology'.

Despite using a multi-method approach (Denzin 1970; Webb *et al.* 1966), the majority of data presented in the thesis are gathered from the dominant technique of conducting semi-structured interviews with leading scientists. Nevertheless, the ***verbal descriptions, explanations*** and statements are often situated and contextualised from the knowledge gained, and information found, in utilising all the methods used in the fieldwork. In combination, these 'other methodological techniques' have helped to produce a more thorough understanding of the social processes at play, and have aided the analytical process. I now continue this chapter with a reflective description of the methodological process.

NEGOTIATING ACCESS AND GATEKEEPERS

Negotiating access to a field requires careful balancing (Hammersley and Atkinson 1995). It was certainly a process that I was acutely aware would need assiduous attention in order to gain access to a group of actors who may not be used to the types of method that I utilise. Acquiring access to respondents involved two levels of negotiation. These two levels of negotiation involved two of the three stages of methodology that made up the study, the site visit (Stage One) and semi-structured interviews with scientists (Stage Two). I begin by describing the process of negotiation for Stage Two of the investigation.

In Stage Two, my aim was to gain access to senior scientists, whom in the majority of cases were academic scientists who were involved in, or affected by, proteomics and bioinformatics work. I began with the notion that this group of actors would be extremely busy and would have limited time to be interviewed by a sociologist. In this respect, I had prepared myself for a series of knock backs. In retrospect, however, it was the very detail that most of my target informants

were academics that I now believe allowed me great access to the field. The fact that the respondents were involved in the creation of knowledge for a living, and understood the notion of research meant that I had a very favourable response from academic scientists working in the field. Although never fully understanding the aim of the research, academic scientists seemed to realise that I too was attempting to create *'scientific'* knowledge from conducting a piece of research.

Many of the initial respondents were found by browsing the World Wide Web (WWW) and typing into *Google* and *Google Scholar* key words such as an area of the UK and the key terms 'proteomics' and 'bioinformatics'. If I found any *hits*, I would then follow this up by reading their biographies and deciding if I should contact them. Other potential respondents were then suggested to me by the initial respondents that I had interviewed. This type of *referral* or *snowballing* process (Vallance 2001), where participants put me in touch with other like-minded scientists was of great help and meant that I quickly established a network of contacts from slightly differing perspectives, but all of whom were working in a similar area.

Once I had targeted a potential respondent, much of the initial correspondence was conducted through electronic mail (e-mail), with any further negotiation performed either on e-mail or on the telephone. In earlier research (Lewis 2006) I have explained how electronic mail is virtually instantaneous and allows potential "participants to receive, deliberate and to respond to [any] questions in their own time" (p5). This flexibility coupled with my inside knowledge of how many hours most academics spend on the computer a week, meant that e-mail seemed the most appropriate form of initial contact. Despite failing to gain a response from half a dozen potential respondents (which I was later told by some of the respondents interviewed, was probably as a result of how busy they were)[42], overall this technique worked remarkably well. Further interaction using e-mail allowed me to negotiate suitable dates, places and times for conducting the interview. Notwithstanding the great flexibility that e-mail provides, perhaps the

[42] In fact, two researchers responded to my request a couple of months after I had finished conducting the research stating that they had only just got round to replying to the e-mail.

most attractive feature was the fact that even after interviews were conducted, a number of participants continued to keep in contact with me via e-mail, often alerting me to workshops, clarifying statements made, inviting me to present at events and adding me to their distribution lists.

The second group of actors with whom I needed to negotiate access, occurred at Stage One of the methodological process. This stage involved a site visit at the European Bioinformatics Institute (EBI) to observe and interview members of the Proteomics Standards Initiative (PSI). Rather than browsing the WWW for key terms, this time negotiating access involved the help of my supervisor who had worked with this group of actors over the last few years. It was at this stage that I began to see the benefits of working on a research topic under a flagship project at a research centre. Invariably those working under the same themed project are researching in similar fields and, because of the closeness of these topics, I believe this allows researchers to share contacts and suggest participants. My supervisor became a key gatekeeper in aiding access to this group by suggesting potential interviewees and helping me frame the initial access-requesting e-mails. This worked advantageously, and I have continued to keep in touch with the six respondents I interviewed at the site visit throughout the duration of the project.

My supervisor was one of two important gatekeepers in the project. The second was also a key interview respondent in the research and has been given the pseudonym, Dr. Campbell. Dr. Campbell was extremely helpful throughout the course of the study. He not only allowed me to interview him twice, but also suggested other potential respondents, set up and arranged an interview with a leading bioinformatician, kept me abreast with any latest developments at his laboratory, and was a key gatekeeper for gaining access to Stage Three of the methodology; a Research Students' Skills Development Programme (RSSDP) course on *Protein Bioinformatics*. Dr. Campbell is one of several participants in the study with whom I have continued to keep in contact, and the relationship I have built up with these respondents has been pivotal to the success of the study.

Dr. Campbell has also been categorised within the study as a core researcher since his main research interest is in proteomics. Those working for the EBI and a

select few whose main interests are with proteomics and standardisation in proteomics have also been classified as core researchers. While those who have additional interests have been labelled as peripheral researchers (Chapter Five).

SAMPLING – WHO TO STUDY?

Implicit in negotiating access with participants and deciding on the type of method to use is the sampling process. When trying to explain my research to some of the scientists I interviewed, I sometimes got asked the question: "so are you like a journalist then?" My original response was: "well similar in some ways I guess, but no not really". I thought it was best not to explain the whole background of the sociology of science until after the interview and only then if they expressed an interest in knowing. Nevertheless, reflecting back on the methodological process, I realised that this question had struck a chord with me, since it remained one that I continued to think over. What really are the differences between sociology and journalism and where do they lie?

In writing this chapter, I believe the answers to the two questions are that there are many differences between sociology and journalism but that the fundamental departures are during the preparation and analysis stages of the process. In terms of preparation this begins with decisions about whom (which actors) to study in the first place and finishes with selecting which data should be used to support analytical arguments. As Flick (1998) explains, the key issue is to be representative from the beginning to the end of the sampling process:

> "The issue of sampling emerges at different points in the research process...In an interview study, it is connected to the decision about which persons to interview (case sampling) and from which groups these should come (sampling groups of cases). Furthermore, it emerges with the decision about which of the interviews should be further treated, i.e. transcribed and interpreted (material sampling). During interpretation of the data, the question again arises with the decision about which parts of a text should be selected for interpretation in general or for particular detailed interpretations (sampling with materials). Finally, it arises when presenting the findings: which cases or parts of text are best used to demonstrate the findings (presentational sampling)?" (p62).

I have used Flick's exemplar as the template of *doing* good qualitative research. As already discussed, the choice[43] of who to interview (case sampling) began with negotiation with a particular core group through one of my supervisors. My supervisor had already been working within this field and with this group and, thus, research evidence suggested that this group were central to conducting research on standardisation within proteomics. Meanwhile, other interviews were negotiated through the other main gatekeeper (Dr. Campbell) and by searching on academic sites on the WWW. In this case I was aware that I needed to get a representative view from (i) scientists working on omic biology, (ii) scientists working on the peripheries of omic biology, and (iii) scientists who were working within more reductionist models. This was required in order to get an overall view of the scientific network. I believe I succeeded in producing this representative sampling frame by recruiting biologists, computer scientists, chemists and mathematicians working in bioinformatics and proteomics. In some cases initial approaches to certain members of the sample population (often the gatekeepers) led to further contact with other respondents and I found this type of snowball effect a very efficient technique when conducting research on community/network based groups such as scientists (see Vallance 2001 for snowballing).

The next stage of the process to which Flick refers is the material sampling stage. In this context, all of the interviews were transcribed and so potential problems of material sampling were overcome. It is perhaps within the context of presentational sampling, however, where one of the biggest differences between a sociological and journalistic account of science lay. Decisions about what to present as empirical material are made in both the collection and interpretation stages of the methodological process (see semi-structured interviews). Furthermore, this works in two opposite directions. As a researcher you may get an innate *feel* of what is the best data, and when these *standout* extracts are confirmed by other respondents' views, you begin to get a *sense* that this data should be included. Or, alternatively you may begin to collect similar type themes/stories/issues from various respondents and decide to *pick out* the most

[43] Obviously you do not have complete freedom of choice since some respondents may decline to be involved in the research.

pertinent extract. The data is then analysed and framed through a sociological perspective, as demonstrated in Chapter Five where the actor-defined buzz word has been translated to the analyst-defined proto-boundary object. This type of analysis differentiates sociology from non-academic journalism. Sociology is more than just (a balanced or biased) reporting of an event, it is also about describing, explaining and analysing the social arena often, but not necessarily, in relation to policy. During the interviews a significant number of respondents expressed an interest in the study and asked if they could have a copy of the completed thesis. If they are still interested and do read the thesis, it may be interesting to gather opinions from the same participants who asked whether my work was a type of journalism, on whether they now have a different perspective of what is involved in social science research.

ETHICAL ISSUES

Confidentiality and anonymity are fundamental when interviewing respondents whose answers and beliefs may have both an impact on their self-identity and how they are perceived by other colleagues. Many of the scientists I interviewed have a considerable status within their communities and are readily identifiable within them. Consequently, a misunderstood extract attributed to them could have the potential to impact on their scientific identity, and damage their scientific standing. This visibility meant that before all the interviews, I provided an informed consent form for respondents to sign if they wished to proceed. The consent form explained that I would use pseudonyms for respondents' names so that no-one was transparently identifiable and also explained that they were under no obligation to answer all the questions. Although using pseudonyms has often been used to protect respondents who may be in vulnerable positions (for example children, the elderly, the sick), I believed it was also the appropriate course of action for this study. The reason behind this was to ensure that any remarks about employers, scientific perspectives or fellow scientists that might be deemed controversial were not attributed to that respondent. However, coding particular organisations or group names has been more complex. It would have been ridiculous to change the names of the organisations such as the EBI and PSI since

the study would have lost all sense of meaning. Thus, even though individuals have been coded certain associations have not (Appendix One).

On a few occasions when respondents wanted to say something that they believed was controversial or not for public distribution, they asked for the audio recorder to be turned off. I found this action very interesting, because despite asking me to turn the recorder off they still wanted to answer the question. Even when researchers with the best intentions attempt to distance themselves from these comments, the researcher often finds that they do influence their research, because if there is one memory that remains vivid after the conclusion of an interview, it is the comments after the respondent says: "I would like you turn off the recorder for this answer". This is an interesting dilemma and poses the question that if the respondent really did not want to answer the question or to reveal anything controversial, why did they not just refuse to answer the question (an option given to them at the beginning of every interview), instead of asking to answer the question unrecorded? I felt that participants in these situations *did* want to give this piece of information to me, but to deliver it in a sanitised way with their names disguised. Consequently this is what I have done on a few occasions in the thesis. I now proceed by giving a description of the methods used.

SITE VISIT/OBSERVATION (STAGE ONE)

The first stage of the fieldwork process involved a site visit to the European Bioinformatics Institute (EBI) based at the Wellcome Trust Genome Campus in Hinxton (Cambridgeshire, UK). The EBI lies in fifty-five acres of parkland that is hidden away off an A-road, and has been home to some of the more prestigious cutting-edge scientists conducting research in the UK (Photo One).

Photo One: The Welcome Trust Genome Campus and surrounding grounds. Photo taken January 2006.

During this one-day visit, I conducted six interviews at the site; four with leading proteomics and bioinformatics researchers involved with the Proteomics Standards Initiative (PSI), and two further interviews with three EBI workers involved in their outreach project. I was also given a tour of the campus by one leading scientist, which included the EBI Building (Photo Two), their computers, the cafeteria, and the surrounding grounds.

Photo Two: The European Bioinformatics Institute (Right). Photo Taken January 2006.

The building itself is relatively new (established 1995), with an airy modern feel, and is in contrast to the old dilapidated Sanger building that is located opposite. On approach to the front of the building you are greeted on the left by a rather imposing large tree (Photo Two) that has subsequently become one of the emblems of the EBI, while through the set of transparent sliding entrance doors was a model in the reception area of the proposed expansion of the building (Photo Three). In some ways this model epitomises the expanding nature of and dependence on, bioinformatics in the biological world (see Chapter Seven)[44].

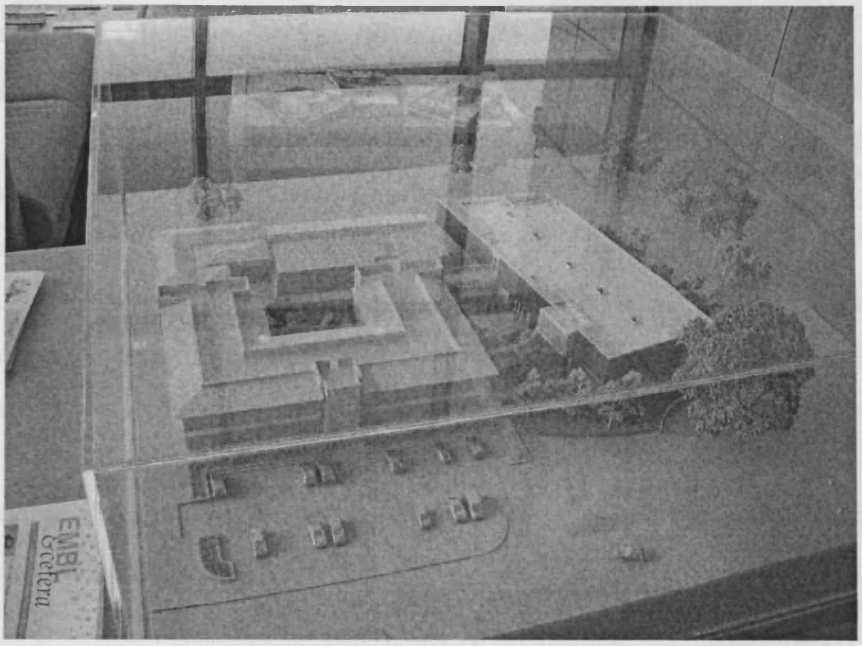

Photo Three: Model Expansion of the EBI (January 2006). Opened in October 2007.

The interviews conducted during the site visit were with the core researchers working in the area of proteomics. Extracts from these interviews are discussed in Chapter Six and the setting equates to a type of micro case-study within the larger project in which I studied (and interviewed) core researchers in their natural locale (the EBI). Observing the setting of the building, the surrounding grounds and the design of the campus aided in *'painting a picture'* of bioinformatics relationship with mainstream biology. The EBI is a new, expanding modern building that symbolises a new, emerging and widening research area, and yet when many people hear the term 'Wellcome Trust Genome

[44] The East Wing was opened on the 23rd October 2007.

Campus' they often think of the older, but more famous Sanger Centre. This view could also be seen as a potential reflection of bioinformatics' relative peripheral status within biology (Chapter Seven). Interview extracts from scientists working at the EBI are displayed throughout the thesis.

SEMI-STRUCTURED INTERVIEWS: THE STRUCTURE AND THE SETTING (STAGE TWO)

The literature on the strengths and weaknesses of semi-structured interviews is extensive (Coffey and Atkinson 1996; Denzin and Lincoln 1994; Flick 2002; Hammersley and Atkinson 1995). According to May (2005): "interviews yield rich insights into people's biographies, experiences, opinions, values, aspirations, attitudes and feelings" (p120). This section will explain the reasons for using semi-structured interviews at Stages One and Two of the research process.

Due to the number of interviews, it was felt that a semi-structured approach would be best to maximise time and content. Semi-structured interview is a technique that sits in between the structured and unstructured interview approach. This methodological approach provides the data for the substantive body of the project and was used to gather opinions and experiences of the social worlds of proteomics and bioinformatics.

During the course of seven months, I conducted thirty-one semi-structured interviews with biomedical scientists based in the UK. Seven of the interviews were conducted with the core proteomics researchers during Stage One of the fieldwork (the site visit to the European Bioinformatics Institute). The other twenty-four interviews were performed at the workplaces of research scientists from five other universities in the UK, and included a mixture of core researchers and more peripheral scientists who were part of the wider scientific or scholarly community. Dr. Campbell (a key gatekeeper in the project) was also interviewed twice during the fieldwork. He was interviewed once at the beginning of the fieldwork and once towards the end.

Interview questions were specific but flexible allowing freedom to diverge from key questions in order to probe interesting answers. Questions explored different themes and a *base* respondent interview scheduled was used. During the fieldwork process, questions were continually added to this *base* schedule and the final version contained over one hundred questions. Some of the questions were specifically relevant to the particular respondent, for example biographical details, while other questions were more universal in their application. The number of questions asked in a particular interview depended on (i) the length of the interview, (ii) the particular respondent and, (iii) their enthusiasm for my research. Interviews lasted on average between forty-five minutes and ninety minutes, and following Fielding's (1988) lead, questions were usually structured by topic: "...they were semi-structured by a thematic guide with probes and invitations to expand on issues raised" (p212). The overall aim was to discover the opinions and experiences of the scientific, technological and social processes involved in proteomics and bioinformatics from the perspectives of scientists directly involved in the areas.

Interviews were conducted at respondents' workplace, or on a few occasions, at their department's café or tea room. The location of the interviews gave me a visual sense of scientific work in action. In many cases this meant having to walk through rabbit warren-like corridors in an endeavour to find the correct room number or laboratory space. The interviews conducted at the scientists' desktop were often in small, dark, enclosed rooms and these were often found in the older parts of the university. In some instances the rooms were shared with other colleagues, but despite this, in the majority of cases I found them a peaceful setting to conduct an interview. Whereas participants' offices were claustrophobic, the interviews conducted in laboratory spaces were often much lighter and spacious. Nevertheless, I found this experience similar to playing cricket next to a motorway, because even though the setting seemed tranquil there was always a hub of activity around me and a constant background noise of the chugging, clanging and murmuring of machines. This background noise at the workplace became an important issue because on five separate occasions I found the setting where I conducted the interview a difficult one to manage.

At the EBI, interviews were held in two different settings. The first was in the open meeting room upstairs, and the second was in the communal café. These spaces were often used as thoroughfares and were home to other EBI workers who tended to use them as spaces to eat, drink, and talk. In both arenas I was immediately aware that they would be difficult places to conduct interviews, but I was assured by respondents that they would not be 'too bad' and that actually there was nowhere else to go. But if these settings were not 'too bad', they were not 'too good' either. It transpired that they were not conducive to uninterrupted interviews. For example, in the meeting room, I was positioned next to a swinging door and had to compete with people walking in and out, while at the café I had to drown out the clamour of lunchtime and mid-afternoon breaks. Despite my initial reservations, and later fears, the recordings were in the most part of decent quality. Nevertheless, if I had been able to choose my own location I would have still preferred to have conducted the interviews in less noisy spaces.

If there is one thing that I have learnt in conducting this research is that it is difficult to *negotiate* spaces when your respondents do not provide any alternatives, and it is a difficult subject to insist on when you are indebted to them for agreeing to be a part of the research in the first place. The result of this meant that similar location issues occurred during other interviews around the UK. On one occasion, when I arrived for an interview, the respondent had completely forgotten that I was coming. I offered to come back at another time but he said that we should conduct it then. It soon became apparent though that he *was* busy with other issues, and half-way through the interview his telephone rung. He apologised, but said he had to answer it. There was about a ten minute gap in the interview while he spoke on the telephone and I sat about three metres away from him in his room feeling as if I were intruding. This was another occasion when I had to do the best with what I had. Other instances included interviewing in a laboratory while competing against telephones, computers, machines and building work outside, and conducting an interview in a shared office where I became acutely aware that I was beginning to disturb my respondent's office colleague, who obviously had no idea that I was expected.

As alluded to earlier, these instances were in the minority and the majority of interviews were conducted in relatively peaceful settings, however, the examples given were the results of not being able to conduct the interview in a neutral setting. Despite my best attempts to *manage* the situations I was still fearful that some of my recordings would not be of good enough quality to transcribe. Luckily my fears were misplaced, but it did mean having to listen to some sections of the interviews on a number of occasions to hear what was said. In most cases, I believe these were occupational hazards of working in a field where you have to interview scientists and begin to immerse yourself within their culture. The problem being the trade-off between doing fieldwork and conducting interviews *in situ* versus performing fieldwork in less '*scientifically*' useful, but more secluded settings.

RSSDP PROTEIN BIOINFORMATICS COURSE (STAGE THREE)

Following an interview with Dr. Campbell, and after encouragement from my supervisors, I decided to embark on a Research Student's Skills Development Programme (RSSDP) course on *Protein Bioinformatics* as the last major methodological technique. The RSSDP is jointly co-ordinated by a number of different graduate schools throughout the university and its remit is to

> "...assist researchers to fulfil the Joint Statement of Skills Training requirements issued by Research Councils and Quality Assurance Agency for Higher Education. The Joint Skills Statement emphasises the importance of generic skills development alongside specific research skills and techniques" (Cryer 2007, p2).

The particular course I attended was organised by the Bioinformatics and Biostatistics Unit (BBU). It was the first year the course had run and its intention was to provide a generic introduction to the emergent field of protein bioinformatics and proteomics for postgraduates and junior staff members. Some students also enrolled on the course as part of an MSc degree, I did not. Instead, I attended five lectures that were focussed specifically on the proteomics section of the course. The lectures were delivered by Dr. Campbell and other colleagues every Friday and Wednesday between May 19th 2006 and June 2nd 2006 and

covered aspects of proteomics and bioinformatics within the context of the technology and the science.

I participated in the course for two main reasons. The first was to use the course as an extra source of scientific information in the areas of proteomics and bioinformatics. The belief was that the course would give me greater background knowledge about the research area that I was researching and would also enable me to gain a greater breadth of scientific understanding (see *interactional expertise*). Attending the five lectures and collecting the accompanying hand-outs proved to be a fruitful exercise is this respect, since it not only gave me confidence that my understanding of the area was of a sufficient level, but it also clarified a number of other issues and introduced me to new questions.

If the primary intention behind attending the course was to extend my knowledge of the scientific literature, the second reason was as part of the fieldwork process. I wanted to attend the course as an observer of how interdisciplinary, emerging post-genomic biology is being taught. Using the traditional observational technique of making *fieldnotes*, I was particularly interested in the translation and transference of emerging omic knowledge from teacher to student (see also Chapter Nine). This type of data collection was also used in various scientific presentations and seminars that I attended.

The group that attended the course was rather small with numbers varying between four and six post-graduate students in some lectures[45]. As expected, the majority of the students were Ph.D. students in bioscience, and I was the only social science student who attended the lectures[46]. The intimate nature of the lectures and the informal environment in which they were conducted meant I was able to get a good grasp of how computers and technologies are revolutionising

[45] I also attended some computer-led seminars in bioinformatics which had significantly more numbers attending.

[46] On one occasion after attending a seminar in the bioscience computer laboratory, I decided to write up my *fieldnotes* when from over my shoulder I was approached by an over attentive computer technician. 'Jamie Lewis', he said. Rather surprised and disconcerted I said 'yes'. 'Social scientist students should not be in here they should be using the computer in the social science department' he responded. It had seemed that the technician had been monitoring the process of my computer activity (we have a specific department led log in process). I explained the situation to the over studious technician and his response was 'well hurry up then'.

the techniques of bioscientists. It was within this context that the course was set up; to extend the skill base of young bioscientists on the cusp of biology and computer science. Even so, the lectures also moved into other areas with the lecture on standardisation practices in proteomics being of particular interest.

In my opinion, this stage of the fieldwork process was vital for a social scientist working in a natural science setting, and I believe it is an excellent source of information for all sociologists working in scientific settings. As stated earlier, the course served in proving an excellent setting to conduct research by focussing on knowledge-transfer, but also in expanding my own skills set as a sociologist of science.

EMAIL

In addition to the major stages of the fieldwork process, I also sent out questions by e-mail to other scientists. These included:

(i) a set of questions to a Dutch scientist I met in a conference,
(ii) email correspondence with a world-leading American scientist, and
(iii) a standard ten-question questionnaire sent to all the directors of the systems biology (SB) campuses in the UK.

This technique had limited success with a number of correspondents failing to reply[47]. Nevertheless, some answers proved interesting and have been included as extracts in the thesis. The second section of this chapter focuses on the role of the researcher in the research setting.

ELITE INTERVIEWING

In this section, I explore and describe some of the experiences I encountered when conducting face-to-face semi-structured interviews with 'elite' scientific respondents. The word 'elite' is *malleably adopted* and the term encompasses a broad range of heterogeneous groups. The American heritage dictionary defines

[47] The three who did reply have been coded and included in Appendix One.

an 'elite' as: "a group or class of persons or a member of such a group or class, enjoying superior intellectual, social and economic status" (American Heritage Dictionary of English Language 2007). This characterisation is consistent with Mills (1956) who views elites as a person or persons who are located within the higher echelons of society. Under this definition, scientific specialists who have a superior knowledge about certain scientific concepts compared to their lay counterparts may be regarded as an 'elite' in that domain. They are at least elite in terms of the notion that they are the leading group of people involved with, and having specialised knowledge of, proteomics and bioinformatics.

Nadel (1956) expands on the ambivalence surrounding the word elite by stating: "there is, in effect a hierarchy among elites, some elites are more elite than others" (p420). In this case, the scientists interviewed in the study may not have been elite in the way a Prime Minister, a Sea Lord or Bill Gates might be considered an elite member of society, but they are elite in the sense that they are the leading researchers in particular areas of expertise. In their article about interviewing elites, Odendahl and Shaw (2002) help to differentiate between these different groups of elites. The biochemical scientists involved in my academic project are categorised under the title of 'professional elites', a rather wide category that includes other professional workers such as lawyers, celebrities and clergymen. This type of elite is based on their hierarchical position of being an expert in a particular field (see Flick 1998, pp91-92 for expert interviews), which in turn, is dependent on their ability and knowledge base. Despite my position as a rather 'inexperienced' sociologist working in a different academic discipline to the respondents I interviewed, many of the respondents were working in a similar working environment (universities). This meant that there was another status imbalance. They were not only elites who had a greater knowledge about the scientific area under study than I, but they were also higher up the academic hierarchy (the setting in which I worked). For instance, I was a Ph.D. student (junior researcher) and they were often Professors or Senior Lecturers (senior researchers).

Stephens (2007) discusses some of the issues Ph.D. researchers must overcome when interviewing research participants in more prestigious positions.

Developing upon the work of Aldridge (1993), who advocated the need to focus on the commonalities between researcher and respondent as well as the differences, Stephens (2007) makes the comparison of a "relatively novice researcher" (p203) interviewing an academic scientist akin to attempting to develop a relationship "mimicking the supervisor/Ph.D. student form" (p208). The focus of this is the importance of rapport building. According to Stephens, building up a relationship with 'the other' and creating an environment in which both members feel comfortable enough to exchange ideas and have a *conversation* is integral when attempting to bridge the status gap. In this scenario I found commonality in the fact that I had a shared academic culture with my respondents. They understood what it was to do a Ph.D. (because most had completed one themselves) and appreciated the notion of research. These were aspects in which I attempted to build camaraderie with respondents.

Spradley (1979) also argues that the establishment of the rapport process is fundamental when conducting any interviews. His four stage model of *introduction, exploration, cooperation* and *participation* is an example of how the researcher can slowly build a relationship with their respondent in order to produce the most conducive setting for information gathering. While, Odendahl and Shaw (2002) believe that the relationship between the interviewer and the elite is further dependent on the setting (see *semi-structured interviews*). Odendahl and Shaw (2002) state that:

> "The dynamics that operate during the interaction are strong and prescient, often constrained by the demands of time and place. The environment where any interview takes place has a bearing on the richness of the data collected" (p304).

This is consistent with my comments on the research setting, in which scientific research participants have a large influence on where the interviews are to be conducted. As such, this adds to their status of being in *their* surroundings and can, as Odendahl and Shaw (2002) suggest, have an impact on the interview dynamics by further widening the status imbalance. Nevertheless, as an interviewer you have to call on all your resources and skills to *manage* the situation, in some instances this means *acting* in an uninformed manner when you

want a participant to elaborate on a point *(exploration)* and other times using your scientific knowledge to show that you can hold a relatively informed scientific conversation *(participation)*. I found that this technique proved extremely fruitful when negotiating Spradley's (1979) four stage model of rapport building with elite respondents.

ACTION RESEARCH

One of the major issues Hammersley (2003) alludes to when reflecting on qualitative methodology is the role of the researcher. In 2002, the Economic and Social Research Council (ESRC) set up and funded the Genomics Network. The network included three centres called the Centre for the Economic and Social Aspects of Genomics (Cesagen), the Centre for Genomics in Society (Egenis), and the Centre for the Social and Economic Research on Innovation in Genomics (Innogen). Today, it also includes a separate ESRC Genomics Policy and Research Forum, colloquially known as the 'Genomics Forum'. The forum's particular role has been to connect social science research conducted on genomics with science policy. According to Webster (2007a), this type of policy work should include both natural scientists and social scientists working together in tandem. In a presentation at the 2007 British Sociological Association (BSA) conference and earlier that year in the 4th International Cesagen conference, Webster stated that he believed social scientists should play an active role in the process of translating scientific work into society. This statement, I argue, is consistent with observations made by McNally (2005) and McNally and Glasner (2005) when inventing the term 'Sociomics'. They use 'Sociomics' to describe what they see as the current two-way traffic of approaches between STS and the omic disciplines; a process that involves the two groups actively collaborating with one another. McNally, Glasner and Wynne (2007) state that:

> "perhaps unsurprisingly, one of the things we have found inside the omic Black Boxes is ourselves. When we examine the omic knowledge-making apparatuses and practices we find social scientists inside the Black Boxes, making a difference to the knowledge that is produced". (p2).

If we are to agree with the statement by McNally *et al.* (2007), it would appear that social scientists have become part of the omic knowledge-producing process

that they are studying. I argue that in some senses, this revelation can be viewed as a type of Action Research.

The development of Action Research in the UK has been heavily linked to educational research. Originally coined by Lewin (1946), the meaning of the term is varied. Despite this, Carr and Kemmis (1986) define Action Research as:

> "...simply a form of self-reflective enquiry undertaken by participants in social situations in order to improve the rationality and justice of their own practices, their understanding of these practices, and the situations in which the practices are carried out" (p162).

In this particular discussion, I too focus on Action Research as self-reflection, but self-reflection as part of potentially participatory research. That is to say, I reflect on two small occurrences in the fieldwork where I began to feel like a *participatory impostor* in my own research, and as a small part of the omic knowledge-producing process that I was studying.

The first and most notable examples of these situations occurred as part of the RSSDP course, during a seminar on proteomics standardisation. Dr. Campbell, the science teacher (Chapter Nine), who knew the nature and content of my research, suggested that I might like to input some of my knowledge of proteomics standards to the rest of the class. It was at this point that I began to reflect on my role as a social scientist conducting a piece of fieldwork on natural science. Was I really in the process of co-constructing knowledge on the subject of standardisation in proteomics, and was I willing to disseminate this to the rest of the class? Was I even actually in any authoritative position to do so in the first place? And perhaps more interestingly, was I beginning to become a respondent/actor in my own fieldwork?

I nervously attempted to gloss over Dr. Campbell's request by stating that I did not really have anything to add to what had been presented on his slides. To my relief, Dr. Campbell did not press me any further on the subject. Nevertheless, at the end of the seminar he did ask me a further private question, which once again led me to scrutinise my position as a social scientist researching in the area of

omic science. On this occasion his question concerned my findings and my relationships with some of the core proteomics researchers who were respondents in the research (Chapter Six). It became apparent to me that Dr. Campbell was wishing to forge stronger links with this group and thought that I may be a good person to start this process. Once again I began to reflect on my position in the study, since this time I really began to feel like a potential science manager who had the opportunity to matchmake two separate scientific groups. I believe this put me in an interesting ethical situation. I felt a strong desire and duty to reciprocate some of the generous help that Dr. Campbell had given me throughout the course of the research and to tell him about the PSI group, and yet at the same time I could not help but wonder if this would *contaminate* the research. Is it really a social scientist's place to matchmake (Chapter Eight) different scientific actors and groups, or could this potentially contaminate any data collected? As Webster (2007b) states when reflecting on science's engagement with policy:

> "maintaining the critical voice of STS here requires continued reflection on the terms with which we enter such domains and the specific discursive spaces they engender, and avoiding the often cynical way in which STS is recruited" (p611).

In my scenario, I needed to balance STS's critical voice with the terms in which I negotiated access with my respondents. I managed to once again avoid Dr. Campbell's question by talking generically around the subject and he has not brought up the subject since. This incident has remained with me though and has left me with a touch of guilt, for I am eternally grateful for all the help Dr. Campbell has given me throughout the study.

The second incident I would like to detail occurred during the site visit to the EBI. At the site, some of the scientists introduced me to the *Toucan* tutorial help tool (Chapter Nine), which is found by clicking on a small Toucan icon on the front of the EBI web-page. Having browsed the page many times before the visit, I found it interesting how I had never come across the tool before. This intrigue remained with me for the rest of the data collection process. During one of the semi-structured interviews, and after having a conversation with one of the respondents about their usage and choice of online databases, I asked whether they had come

across the EBI's Toucan tutorial. I repeated this question in a further half a dozen interviews and on each occasion the participant replied explaining that they do use the EBI online warehouse but had never seen the tutorial before. Furthermore they also remarked on how useful it would have been if they had known about it since some had found the website difficult to navigate.

To begin with I thought nothing particularly out of the ordinary of this line of questioning, because my intention was to gather information on scientists' views of the user-friendliness of certain omic online databases. In fact, I used the question as an ice-breaker in the interview. Nevertheless, when I came to reflect upon some of the methodological issues of the study, I realised that this was actually another type of participatory action research whereby I was matchmaking separate scientific groups. Was there any real difference between the Dr. Campbell situation and this one? For in effect, I was acting as a mediator advertising an EBI web tool to peripheral actors/users. Moreover in this instance, I was freely instigating the process by asking if they used the Toucan tutorial, and then directing them towards the icon on their computer screen.

The two examples I have described, I believe, question the social scientist's role as a *stranger* in the field. They also pose the under-researched question of whether social scientists working in the field of science and medicine have a history of becoming part of the research process, and if indeed they have, whether they should do or not. I suggest that other social scientists, for example criminologists doing research on crime and justice, do not have a history of matchmaking criminal groups with one another in the same way that I was beginning to matchmake scientific groups. Of course they have a history of immersing themselves within the culture they study, and possibly of directing crime victims to particular organisations, but perhaps they do not have a history of unintentionally influencing the *ideas* of that culture in the same manner. Consequently, I am left wondering if our research questions are different from those of criminologists or whether our research settings are so different that they promote different ethnographic roles for the social scientist. For example, a number of STS academics have started publishing in the journals of the

communities they are studying. These are interesting reflexive questions that I believe require further research and interrogation.

INTERACTIONAL EXPERTISE: THE SOCIOLOGIST OF SCIENCE

The most apprehensive, and yet exciting section of the project was embarking on the fieldwork process. My disciplinary background was as a sociologist who had worked in the area of medical sociology and on issues of futures and risk. I had never conducted any science studies work however, and furthermore, I had not studied natural science since GCSE level (approximately ten years previous). Consequently, one concern I had throughout the project was the technical nature of the topic being studied and whether I would have sufficient transferable expertise to conduct the research. Geesink (2006) also had a similar concern when conducting her doctorate stating: "although earlier research had made me familiar with the 'science' behind tissue engineering, the amount and complexity of techno-scientific and clinical data was at times daunting" (p99).

On entering the field, I had prepared myself well, reading the appropriate scientific literature, reading the applicable sociology of science literature and enrolling on the RSSDP protein bioinformatics course, because as Geesink (2006) states: "I had to know the basics" (p99). Nonetheless, I still felt a high degree of anxiety. In this section of the methodological reflection chapter, I focus on reflexivity and use the interactional expertise model (Collins 2004b, Collins and Evans 2002 and Collins *et al.* 2006) to describe and explain both my journey through the interview process, and the new skills I acquired as both a science interviewer and a sociologist of science.

Collins *et al.* (2006) describe how it is possible to acquire interactional expertise. They state that:

> "Interactional expertise' is developed through linguistic interaction without full scale practical immersion in a culture. Interactional expertise is the medium of communication in peer review in science, in review committees, and in interdisciplinary projects. It is also the medium of

specialist journalists and of interpretative methods in the social sciences" (p656).

Essentially, the term interactional expertise describes the ability to immerse oneself into the linguistic culture of a particular specialism without necessarily learning to practice their skills. Or to put it another way it is "the ability to converse expertly about a practical skill or expertise but without being able to practice it" (Collins 2004b, p125). It is a skill that sociologists have been required to learn and refine when conducting their research, since most sociology is the sociology of a particular area. That is to say that very few sociologists do the sociology of sociology; instead they study areas of social life which require them to converse in the language of the particular social group[48]. It is also a skill that is increasingly being desired as we move/have moved into an era where interdisciplinary research is rewarded. This skill is then acquired through linguistic socialisation in a culture, which may take the form of observations or interviews for example.

According to Collins (2004c), a key indicator of someone actually acquiring the skill of interactional expertise is when interviews with respondents turn into the kind of conversations with respondents that Stephens (2007) advocates. What Collins means by this statement is that when the sociologist has gained a grasp of the field so that they can interact in a free-flowing conversation with an expert on a particular subject, they have acquired the skill of interacting in that particular expertise. He even states that the sociologist's level of interaction may even rise to a state so that the conversation may be of benefit to both parties; the sociologist interviewer, and the expert respondent (although this does bring up some of the issues that I described in the Action Research section)[49]. During the journey of the thirty-one interviews I conducted, I certainly felt a dramatic change in both the structure and the quality of the interviews. I suggest this was due to three reasons.

[48] Interestingly Bloor (1976) believes that we should be able to do the sociology of sociology.
[49] In the case of my research this started to emerge when respondents began to ask me about the work of other groups I had studied.

The first was that the more interviews I conducted the more comfortable I felt performing the research, and with this my own confidence increased. Interviewing Professors, Senior Lecturers and leading scientists as a doctoral student is never easy (see *Elite Interviewing*), but with practice and experience came a greater self-belief both in the research and my ability to perform it. The second reason was that through practice my skills as an interviewer grew. Balancing the act of asking questions, listening to the answers, and getting ready for the next question was at first something akin to a juggling act. But later this became smoother, more natural and less forced. Finally, my specialist knowledge and understanding of the particular research area increased.

From the fieldwork journey I experienced, I would argue that gaining interactional expertise as a sociologist of science interviewer means being able to not only learn new knowledge about a research area, but also involves acquiring the abilities of communicating and conversing with others in specialist conversations. That is to say that a researcher acquiring the skill of interactional expertise as a sociologist interviewing for fieldwork must not only learn the concepts of the research, but must also acquire the skills of interviewing and communicating. This is what Collins and Evans (2007) might term '*interactive ability*'. Perhaps the event that signalled my acquisition of the interactional expertise skill was when, after roughly ten interviews, a number of respondents begun to express their enjoyment and interest in the '*conversation*' we were having. There was also a sense of achievement on my behalf. I distinctly remember coming home from interviews thinking that they had '*gone really well*'. Nonetheless, even if you do acquire interactional expertise this is not to say that you will not have another bad interview, since there is no concrete boundary between being a good interviewer and a bad interviewer. Even those people at the top of their professions have good and bad days, and this is no different for interviewing: some interviews are simply just better than others. The quality of a particular interview may be a result of the rapport you have built up with a particular participant or even something as seemingly trivial as the time of day it was conducted. Nevertheless, at this stage of the fieldwork process there was a continual sense of satisfaction after each interview that I had begun to have

conversations with participants, and was *discussing* scientific issues with respondents (see Rubin and Rubin 1995 for discussion of guided conversations).

One final remark about the reflexive process of interactional expertise is to state that there are different levels of interactional expertise, much in the same way that there are different levels of professional football players (e.g. premier league, championship or conference). There are those that are excellent interactional experts and there are those who are mediocre interactional experts. Also in the same way that professional footballers practice their skills to get better, so interactional experts can do the same. Taking this viewpoint, I still learn new facts and gain important information about proteomics and bioinformatics everyday, reflecting and recognising that I am still a relatively junior researcher. Consequently, the methodological process within a Ph.D. does not finish until the thesis is handed in. To illustrate this, I now lead into the analysis component of the research.

ANALYSIS

A clear benefit of conducting your own research project from start to finish, as opposed to being part of a group researching on a project, is that the researcher is involved in the whole research process. Researching in a team means that work invariably needs to be shared, and meaning and analysis is negotiated between the research team (Lingard *et al.* 2007). Conducting your own research, however, should mean that you are familiar with all your data and that this type of negotiation of meaning is not required. I argue that this is important, because it is actually during the interview stage that preliminary analysis begins. As the interviewer who is using the semi-structured interview technique you are, by and large, in control of the setting, since *you* can ask the respondent to elaborate on the issues that *you* feel are interesting and worthy of more discussion. By focussing on certain topics and asking particular questions this amounts, in effect, to a type of preliminary analysis of the contents that *you* (the analyst) find the most interesting. When reflecting on the methodology used in the study, I have truly begun to understand the dynamic interplay between the self (the researcher/interviewer/analyst) and the significant other (the respondent) in

fieldwork encounters, and have recognised how the interviewer can gently orchestrate the ways in which the conversation flows (Atkinson 2006, Scott 2007).

If tentative analysis began at the preliminary stage, the substantial process of in-depth analysis gathered momentum when all the interviews were transcribed. This is a process I now describe. To begin with the data were analysed in-depth by content for emergent themes (Weber 1990). Examples included 'standardisation', 'scientific learning' and 'uses of technology'. These were then coded more specifically into categories (Strauss 1987). Coffey and Atkinson (1996) expand upon the link between concepts and data. They state:

> "Many analyses of qualitative data begin with the identification of key themes and patterns. This, in turn, often depends on the processes of coding data. The segmenting and coding of data are often taken-for-granted parts of the qualitative research process. All researchers need to be able to organize, manage, and retrieve the most meaningful bits of our data" (p26).

It is at this juncture that conducting your own fieldwork becomes beneficial. If you have been involved in all facets of the research, conducting and transcribing all the interviews, then you may *know* your data better, and thus have a greater *feel* for the "most meaningful (and interesting) bits". In terms of this particular research, the predominant and recurrent themes of interest were debates around expertise, boundary classifications, the role of standardisation, craft in science, and emergent knowledge-transfer. The most striking extracts of data are presented in the five data chapters (Chapters Five to Nine) as representative of those themes, and as support for the arguments made.

The coding process began by transcribing all the thirty-one interviews and then reading and listening to them in their entirety. The Microsoft Word files on which these extracts were transcribed were then converted in to Rich Text Files (RTF), and imported into the NVivo qualitative software package. I had originally decided on using the Atlas Ti qualitative package and attended a training course in Guildford. But after completing a similar course for the NVivo package, I believed that for this particular project NVivo was the more user-

friendly programme. Of course NVivo does not *do* any of the analysis. It is, instead, a type of data management, organisational and pictorial tool. Nonetheless, it is a platform where you can *transform* data, *find* emergent themes and *code* by creating trees and networks of trees. Examples of the codes/trees created included the themes 'standards' and 'knowledge'. These were then analysed further for a more narrow analytical category, for example in the standards section there were the codes 'core creation of', 'local effect of', 'implementation of', 'boxed' and 'dominant', while in the knowledge section these included 'lack of knowledge', 'knowledge-transfer', and 'new emergent knowledge. Simultaneously, I also created one hundred and six 'free tree nodes', which included the themes 'authorship', 'boundaries', communication', 'dissemination' and 'expertise'. These themes were created by combining words from the science and technology literature (the STS analyst language) and words from the uncategorised participant language. The themes were translated into nine potential chapter headings including 'boundaries of bioinformatics', 'boundaries of proteomics', 'configuring the scientist' and 'social networks and interdisciplinary research'. It was from these initial chapter suggestions that I collapsed the outline of the Ph.D. into the more manageable five data chapters (Chapters Five to Nine). Without the use of qualitative software platforms such as NVivo this type of in-depth analysis would have been more difficult in the timeframe given to conduct a Ph.D.[50].

CONCLUSION

In this chapter, I have *reflected* on the distinctly qualitative methodological approach undertaken in this project. The data presented in Chapters Five to Nine reflect this *modus operandi*. The chapter has also made explicit my role as the researcher and the gradual development I have made when attempting to cross the social/natural science boundary. This is explicit within the reflexive discussion around the development of my (the researcher's) increased knowledge of the scientific areas of proteomics and bioinformatics. From an *outsider* of science, I

[50] A final comment to make about NVivo is that by creating networks of trees the qualitative software package also helped to visualise connections between what were at first viewed as heterogeneous, unconnected topics.

developed into a type of inside outsider (Reiner 2000) or quasi insider (Emmerson, Fretz and Shaw 2000) of science by acquiring the skill of interactional expert. There is a belief within social science that being a stranger in a new research field has a number of advantages. However, I maintain that there comes a time when you need to have a good basic grounding of the knowledge of an area (for example to be able to conduct interviews at a sufficiently high level). In fact, I have come to the conclusion that the more knowledgeable you are about a specific research area then the better placed you are to conduct a piece of sociological research on it[51], thus explaining my focus on interactional expertise. Nevertheless, knowledge *per se* does not determine your position as an outsider or an insider or even an outside insider or an inside outsider, other events including access to the site and even luck play a large part too. I do believe though that to be able to continue to write with a critical voice then being situated in one of the last two positions (inside outsider or outside insider) is of most value.

To conclude, (as has been explained earlier in the chapter), the main method used in the project has been semi-structured interviews. Recently, Atkinson, Coffey and Delamont (2003) have been quite critical of the over-dependence of this research technique by some colleagues stating that:

> "we think that too many of our contemporaries and younger scholars turn to the research interview as an easier alternative to the harder work of prolonged immersion in a social world...we also have to recognize the forms of talk...are themselves examples of social action" (pp116-117).

Their argument is of course very difficult to dispute, however, the choice of using semi-structured interviews in this study was rather forced on the researcher. This is because the participants in this study are part of a network rather than a local community or group (Chapter One). It was even suggested that in some cases there is no substantive proteomics activity to be found; instead, only proteomics talk (Chapters Two and Five). It is precisely the way that participants talk about modern science and the great aspirations of omic biology which drew me to this project and to the analysis (see Chapters Five to Nine). Consequently, it would

[51] With the added caveat that as long as the researcher is able to keep a distance between him/her-self and the group they are studying.

have made it incredibly difficult and unrepresentative if I had restricted myself to conducting an ethnography of the PSI. Instead, I wanted to get an understanding of the level of development of proteomics and bioinformatics within different organisations in the UK. I begin this quest in Chapter Five where I present data gathered from both *core* and *peripheral* proteomics workers (those involved in the PSI[52]).

[52] Due to the content of chapter, and as a way of helping the reader, I differentiate core researchers from peripheral researchers in Chapter Five.

CHAPTER FIVE:

BEYOND BOUNDARIES: PERFORMING THE PROMISE OF PROTEOMICS

INTRODUCTION

This chapter reflects upon how scientific knowledge has historically been divided and classified into autonomous specialised disciplines. It continues by illustrating how new *hybrid*[53] research areas have regularly developed between two or more of the existing traditional research areas; a prime example of which was the emergence of biochemistry around the turn of the twentieth century as a mixture of biology and chemistry. Applying this boundary model of scientific practice and reflecting on the past, present and possible future, the chapter explores the emergence and potential stabilisation of proteomics through the opinions and accounts of a variety of actors. Using two heterogeneous examples, I begin by introducing the notion that the boundary is a *social construction*, and conclude by discussing the consequences of *imagined* boundaries in relation to proteomics. I also introduce to the literature the notion of the *proto-boundary object*, which I suggest resides in a *phase zero* of scientific development and explore the relationship that this has on determining a scientist's identity and their potential funding possibilities. By including both core researchers and peripheral actors' extracts, the chapter reveals that there are potentially many stories to be told about the emergence of proteomics.

ACCOUNT ONE

> **George Jung:** "Your honour, I'd like to say a few words to the court if I may...Well, in all honesty, I don't feel what I've done is a crime. And I think it's illogical, and irresponsible for you to sentence me to prison. Because, when you think about it, what did I do? I crossed an imaginary line with a bunch of plants..."
>
> **Judge:** "Yeah, gosh, you know your concepts are really interesting, Mr. Jung."

[53] Hybrid in the sense that the new research area is composed of the scholarly backgrounds of two autonomous traditional disciplines.

George Jung: "Thank you."

Judge: "Unfortunately for you, the line you crossed was real, and the plants you brought with you were illegal..."

The above extract was taken from the film, 'Blow' (Dir. Demme), screened in 2001. The conversation is set in a Chicago courtroom where George Jung, an international drug baron, attempts to defend his actions of smuggling marijuana to the High Court judge. His defence is based on his belief that the plants are natural resources taken from the ground and that the international borders are just artificial social constructions. The judge's response to his plea is that unfortunately for Mr Jung, the plants he smuggled are classified by the court as illegal, and that those imagined border lines have real consequences. Herein lays the nub of the judge's retort, and the root of the chapter: that imagined socially constructed boundaries do not need to be physical to have very real political and social effects on society.

ACCOUNT TWO

During the inaugural *'Science Wars'* meeting in 1994 at Loughborough University, an audience member questioned whether the vehement disagreement between the sociologist Harry Collins and the developmental biologist Lewis Wolpert was really a debate about funding. They questioned whether the whole *'science wars'* conflict was, in effect, natural science academics just defending their scientific funding boundaries from perceived STS encroachment. This was dismissed out of hand by Lewis Wolpert who stated that his funding comes from completely different sources to STS academics. Interestingly, the fact that Wolpert claimed he gets funding from different suppliers to academics such as Collins marked another imagined boundary; the boundary between the natural and social sciences (Gieryn 1999). Despite Wolpert's condemnation of this claim, the issue of boundary demarcation in relation to funding arose once more in the meeting as a central concept in scientific practice. Implicit in the discussion was that learning the skill of gaining funding is a significant part of a scientist's performance and identity since without funding there is simply no professionalised science. Herein lays the second fundamental basis of the

chapter: that contestations of boundary work are often implicit or explicit attempts by scientists to gain funding.

ORGANISING KNOWLEDGE: CONSTRUCTION OF BOUNDARIES WITHIN SCIENTIFIC PRACTICE

Science as an organised form of knowledge and as a professional ideology is divided into distinctive and autonomous research areas, the inner contents of which are bordered off by constructed *imagined* disciplinary boundaries. It is organised in this manner so that science as a specialised type of methodology can flourish. The cordoning off of specialised research areas help nurture, among many other things, new technologies, new methodologies, new communities and new disciplines. In this regard, although these boundaries, (like the film example illustrated) are socially constructed or, as the character George Jung phrases it, imagined in form, they do have *real* significances. The demarcation of scientific practice in this way generates synthetically[54] constructed social spaces that

(i) allow scientists to create new forms of knowledge under the comfort, and tension, of being a legitimate *'expert'* in that particular area,

(ii) justify the huge expense involved in purchasing resources and technologies under the haven of specialisation, and

(iii) help form expert collaborations that drive the future of the research.

This all suggests that imagined boundaries need to be carefully constructed for science to progress in a functional, efficient and ordered manner.

This modern boundary map of science can potentially be traced back to the seventeenth and eighteenth centuries, and the period known as the enlightenment era. According to Israel (2006) this was a revolutionary period within natural philosophy. However, and consistent with the work of Shapin (1996), Israel argues that this was a philosophical revolution rather than a scientific one. Despite Merton's (1970) work classifying the late seventeenth century activities of the Royal Society of London into specific fields of interest, such as philosophy, formal sciences, physical sciences, biological sciences and cultural

[54] In using the term synthetic I am describing how disciplinary boundaries have been socially constructed so as to order scientific knowledge.

anthropological sciences, there is a general consensus that cultural and scientific practices of that era were often lumped together in the more universal categories such as *science*, *natural philosophy* or *experimentalism*. Moreover, Shapin (1996) sounds a warning when telling stories that trace the emergence and development of scientific activities. Shapin (1996) argues that:

> "the past is not transformed into the modern world at any single moment: we should not be surprised to find that seventeenth century scientific practitioners often had about them as much of the ancient as the modern...the people, the thoughts and the practices we tell stories about as ancestors...always reflect some present-day interest" (p7).

Consequently it is not until the nineteenth century that we can really begin to see the development of fledgling present day scientific research areas. During this period segregated and autonomous disciplines, such as Darwinian evolutionary biology, began to emerge from original umbrella terms such as natural philosophy. This new nineteenth century map attempted, and indeed succeeded, in defining the individual, specialised subject matter and methods in hand. These demarcations were much more like the historical ancestors of some of the professionalised disciplines that we know today; such as biology, physics, chemistry, psychology or sociology. New and different forms of knowledge were initiated, widening the scope of scientific activity under study, and these were separated from their counterparts by imagined and constructed boundaries. Contained within each autonomous boundary were all the relevant scientific practices of that particular domain. The result of this was, rather than a group of people being known by the rather generic terms such as the *philosophes*, producers of knowledge began to become known by their boundary marker. For example they became known as the biologist, or the chemist, or even the sociologist, with the term scientist only invented in the nineteenth century and only routinely used in the twentieth (Shapin 1996). Although these were still different from the professionalised and sanitised research fields we find in science today they were, however, much more closely related. Simply stated, this type of boundary demarcation implied that all that was contained inside the boundary of *biology* was biological practice, and all that was contained inside the boundary of *physics* was physical practice, and moreover these two scientific disciplines were

epistemologically and ontologically distinct from each other. In fact, Gould (1989) remarks that actually the two areas of research had very little in common with one another to begin with. An example of this sanitised map of science is illustrated in Figure 5.1.

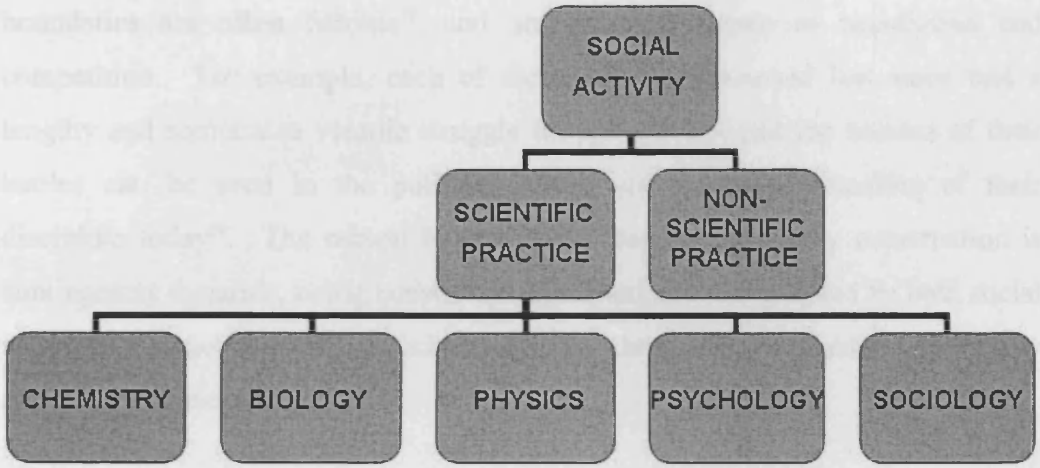

Figure 5.1: Traditional Boundary Map of Science[55].

Figure 5.1 represents the model of science described above. Using a number of disciplinary examples it visualises the synthetic constructed boundaries contained within science that distinguishes one intellectual activity from another.

In reality I believe this model has some problems. I maintain that Figure 5.1 could be called a science sand-castle; a model of scientific activity built upon a foundation of shifting sands. I have called it a sand-castle because in the first instance I have set up the model in order to knock it down as a type of *strawman*[56], and in the second, I argue that 'the model' is continually changing shape and form. This is consistent with Gieryn (1983), who might argue that the type of model of science portrayed in Figure 5.1, although perfectly justifiable in some circumstances by some people at some time, should not be viewed as a fixed model of science. This is because Gieryn (1983) believes that boundaries

[55] This diagram is obviously only a selection of the wide variety of scholarly disciplines that are found within scientific practice, and it is also important to note that non-scientific practices are often prevalent within scientific disciplines.

[56] The model presented is a sanitised, retrospective, rigid representation of science. As such, I set it up as a type of strawman argument in which I argue that scientific disciplines are much more flexible than Diagram 5.1 suggests.

are not fixed entities. Scientific boundaries of the 'natural world' vary over time and change between societies. For example, it may be argued that in the UK (Lemaine *et al.* 1976) and in the US (Gieryn 1999) the boundaries between the natural sciences and the social sciences are marked by a distinctive border, while in French society this is less clear. Subsequently, Gieryn (1983) argues that boundaries are often flexible[57], and are frequently open to negotiation and competition. For example, each of the territories discussed has since had a lengthy and sometimes volatile struggle for recognition, and the success of their battles can be seen in the political, social and economic standing of their discipline today[58]. The reason behind this is because boundary construction is contingently dynamic, being constantly negotiated and renegotiated by both social actors and socio-political structures. Gieryn elaborates on science's flexibility and non-permanency:

> "Science is no single thing: its boundaries are drawn and redrawn in flexible, historically changing and sometimes ambiguous ways" (Gieryn 1983, p781).

Referring back to Figure 5.1 (p114), the rectangle on the top row represents society and all the social activities that can be found within its boundary. In the second row this category is split into two further categories called scientific practice and non-scientific practice[59]. Although not shown, non-scientific practices may include activities such as religion or sport[60], and this boundary is demarcated from scientific practices. What are also illustrated are five examples of autonomous research areas that can be found within the boundary of scientific practice (i.e. chemistry, biology, and physics: the natural sciences, and psychology and sociology: the social sciences). If, as Gieryn (1983) states, the boundary of science is continually changing form, it is reasonable to suggest that what is contained within it will and has continued to change form too. This is most apparent and best demonstrated in instances where, over time, some of the

[57] For example, physical practices can be found in biological experiments such as mass spectrometry.
[58] Often it is those research areas that are the most transparent, and that open themselves up to the most criticism that have tended to become recognised as the most scientific.
[59] It is recognised that there are many non-scientific practises found in scientific practice.
[60] However, lest we forget that non scientific practices can *seemingly* merge with scientific practices. Examples include sport science and the church of scientology.

boundaries have been permeated and/or amalgamated to create merged boundaries, sub-disciplines or new disciplines. Let us take the five scientific disciplines already mentioned and let us assume they are accepted scientific practices. Taking into account Gieryn's view, a new diagram of scientific practice can be drawn to illustrate what has happened to some of these boundaries over time (Figure 5.2).

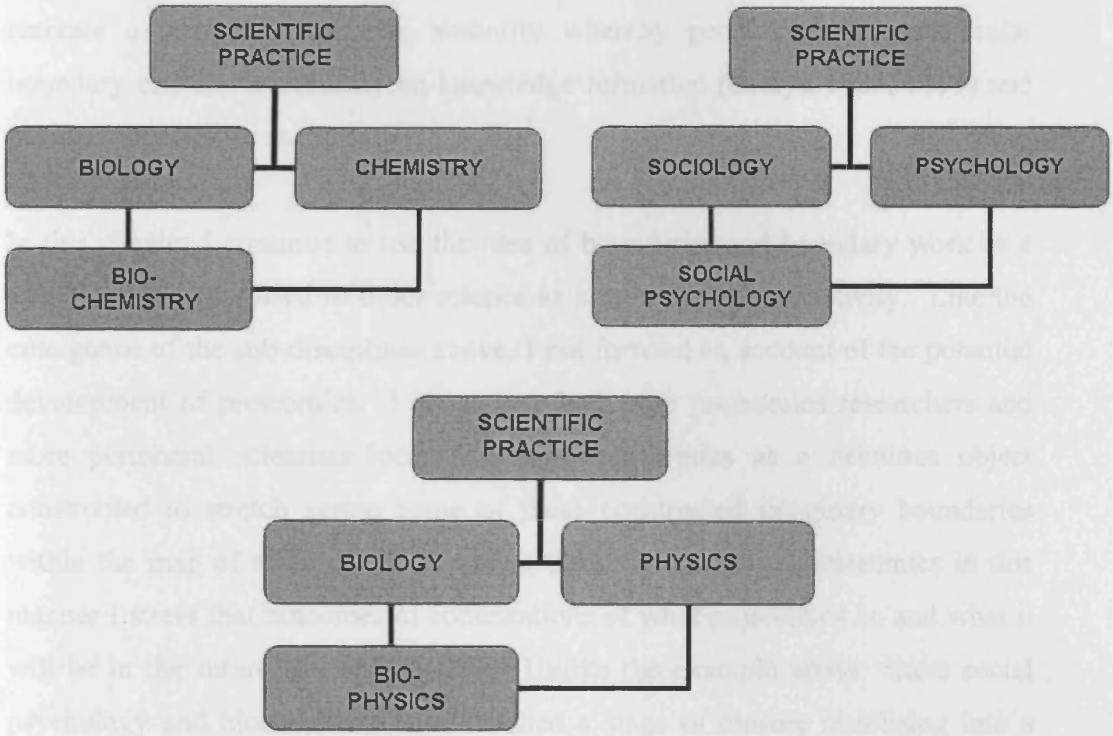

Figure 5.2: Emergence of Sub-Disciplines.

As a result of scientific proliferation and subsequent permeations of original autonomous boundaries, new sub-disciplines or hybrid research areas have emerged over time. In Figure 5.2, three models are produced in order to illustrate the development of biochemistry (a combination of biology and chemistry), biophysics (a combination of biology and physics) and social psychology (a combination of sociology and psychology). Other variations could also have been highlighted such as physical chemistry; the application of physics to chemical systems, or more recently chemical biology. In these interdisciplinary areas the unambiguous boundaries between scientific disciplines in the first diagram such as biology and physics become less apparent with the emergence of biophysics. This in turn has real consequences on scientific practice, since not only are the

scientific practices contested intellectually by different actors, but also actors attempt to claim territories that have real effects on issues of modern scientific funding and expertise. For example, contests may emerge regarding who is the expert in the area of biophysics: is it the biologist or is it the physicist? Modern, professionalised science is inherently competitive and these competitive struggles between different groups of scientists re-emphasise the value and utility of the demarcated discipline. In this regard these boundary formations create and recreate a sense of epistemic authority whereby people inside a particular boundary can claim authority on knowledge formation (Gieryn 1983, 1999) and position themselves as the expert.

In this chapter I continue to use the idea of boundaries and boundary work as a social construction used to order science as a professionalised activity. Like the emergence of the sub-disciplines above, I put forward an account of the potential development of proteomics. I show how both core proteomics researchers and more peripheral scientists locate the term proteomics as a nebulous object constructed to stretch across some of these constructed imaginary boundaries within the map of science. In exploring the development of proteomics in this manner I stress that outcomes of contestations of what proteomics is, and what it will be in the future, are still unclear. Unlike the example above where social psychology and biochemistry have reached a stage of closure stabilising into a standardised package (Fujimura 1992), I display proteomics as a proto-boundary object: a research area which has not yet stabilised to the level of a standardised package but with which heterogeneous actors align themselves. Likewise, by focussing on funding and expertise, I also show how competing scientific actors endeavour to cross the constructed boundaries contained within science as *obligatory points of passage* (Callon 1986) in order to gain funding. Furthermore, I explore the concept of the 'buzz-word' as a rhetorical device that can influence scientific actors to proclaim their work as congruous to other actors in characterisation, but which in practice, may be theoretically distinctive. Wolpe and McGee (2001) remark how in public policy debates about stem cells:

> "the first battle is often a struggle about definitions, and the winning side is usually the one most able to capture rhetorical primacy by having its

> definitions of the situation accepted as the taken-for-granted landscape on which the rest of the game must be staged" (p185).

Within this chapter I argue that the same is true about the emergence of proteomics. Crucially the chapter explores how actors and research groups try to establish their area of primacy through the *malleable adoption* of a new rhetorical artifice (in this case proteomics), while not having to change their existing established scientific identity.

THE BOUNDARY OBJECT

The social constructionist position of science I have portrayed here can be seen in the work of Gieryn (1983, 1999) when writing about the demarcation of natural science and social science, and Hedgecoe (2003) when discussing the influence that terminologies have (or do not have) on the future trajectories of emergent disciplines. Both authors have been interested in the effects that boundaries have on ordering knowledge, claiming territories, directing research and changing identities. It is this last effect (changing identities) that Bowker and Star (2000) explore in their work on classifications and their consequences. Bowker and Star demonstrate how constructed boundaries classify people in many ways, and illustrate some of the ways in which those classifications heavily alter that person's self-identity. They give the example of apartheid in South Africa and show how the dichotomous classification system of Black or White had severe consequential impacts on certain individuals. This classification, often mediated by its social impact, had an enormous effect on the individual's self-identity, and once again illustrates the real consequences of classificatory boundaries. Nonetheless, it is in earlier work with Griesemer on boundaries and classifications that Star coined the term 'boundary object'; an object that can transcend heterogeneous identities and cross boundaries. According to Star and Griesemer (1989) boundary objects are:

> "...scientific objects which both inhabit several intersecting worlds and satisfy the informational requirements of each. Boundary objects are both plastic enough to adapt to local needs and constraints of the several parties employing them, yet robust enough to maintain a common identity across sites" (p505).

In their definition of boundary objects, Star and Griesemer (1989) refer to objects as material things. An excellent example of this would be a map; an object that is able to traverse the social worlds of geographers and walkers for example[61]. In fitting with the rest of the chapter, however, I would like to objectify any boundary permeation as an object. Under this new definition, instead of a classification labelling an actor and consequently changing their self-identity, a boundary object could be interpreted as something that is malleable enough to be permeated by an actor under the freedom to either reconstruct their identity or to continue with their already established identity. At the same time, the object could still be strong enough to integrate the mixed identities. For example, if we accept the two separate classifications of student and worker, an example of this could be a university student who begins to work for extra money. As a (hypothetical) student-worker the individual would not have to conform to the restrictions of their new identity. Instead it would be their choice as to whether they wanted to (i) continue with their traditional identity (student), (ii) change their traditional identity (student) to the 'new' identity (student-worker), or (iii) balance both identities (student and student-worker) so that they co-exist. Consequently a boundary object, under this new re-definition, is not just restricted to being an object that travels between different social worlds picking ideas up from them all. It could also be a classification such as the fictional worker-student example, or it could be a constructed scientific boundary such as proteomics (Figure 5.3).

[61] In fact classifications can be boundary objects; concepts that are able to span different contexts while being operated differently in those settings.

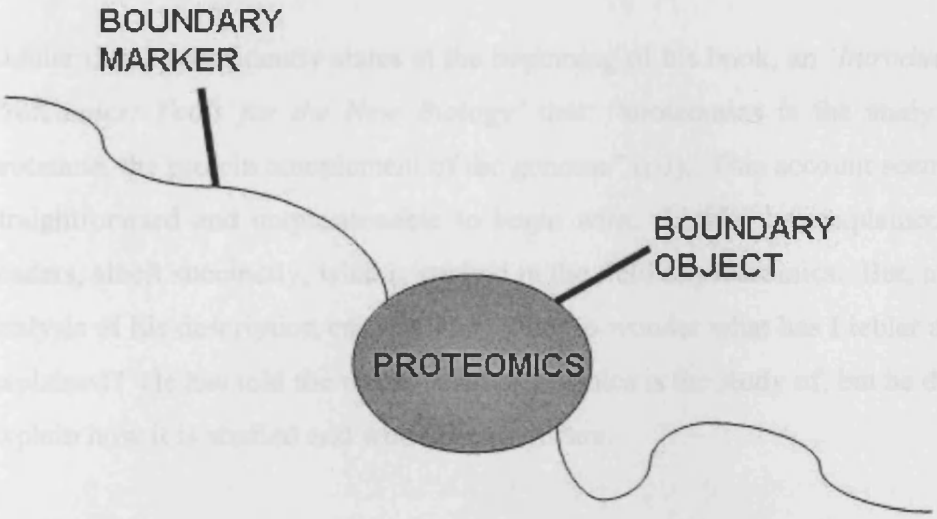

Figure 5.3: Proteomics as a Boundary Object.

The inherent problem of boundaries and the classifications that they embody are that they often become black-boxed. Over time, they can become concealed, forgotten, taken-for-granted and often difficult to cross. The opening of these boundaries, through examinations of boundary objects, however, often tends to reveal interesting modes of translation, interdisciplinary work, and contests of expertise between and within the boundaries. In this chapter, by using extracts from various respondents who define proteomics differently, I refer to proteomics as an interdisciplinary *hybrid* research area (similar to those portrayed in Figure 5.2), but one which has not yet reached the status of a boundary object. Instead, I illustrate how proteomics is a *proto-boundary object*, one which is flexibly used by different actors so that it is elastic enough to inhabit several intersecting worlds, but not yet robust enough to be regarded as a stabilised boundary object. It would seem then that the term proteomics has been able to attract and mobilise numerous experts from distinct, and sometimes competing, disciplines, who have then *malleably adopted* the term to align with their own work. This has meant that any consensus of what the term proteomics entails has become blurred since proteomics tends to mean different things to different actors.

PLACING PROTEOMICS: WHAT IS IT?

> (Laughing) "Do you (the interviewer) want to define proteomics for me?"
> **[Dr. Dennis: Lecturer in Genetics and Molecular Biology Research (Peripheral)]**

Liebler (2002) confidently states at the beginning of his book, an *'Introduction to Proteomics: Tools for the New Biology'* that: "proteomics is the study of the proteome, the protein complement of the genome" (p3). This account seems quite straightforward and unquestionable to begin with. Liebler has explained to his readers, albeit succinctly, what is studied in the field of proteomics. But, a deeper analysis of his description can lead the reader to wonder what has Liebler actually explained? He has told the reader what proteomics is the study of, but he does not explain how it is studied and who the experts are.

Later in the book he continues to describe some of the techniques and technologies used within the boundary of proteomics, such as mass spectrometry (MS) and gel electrophoresis (GE). It then seems sensible for the reader to deduce that proteomics is a discipline or research field; that it is a constructed boundary that may have the very real consequence of producing journals, professional roles and technologies (such as MS). Using the idea of proteomics as a discipline or scientific research area, Figure 5.4 can be drawn to locate proteomics more globally.

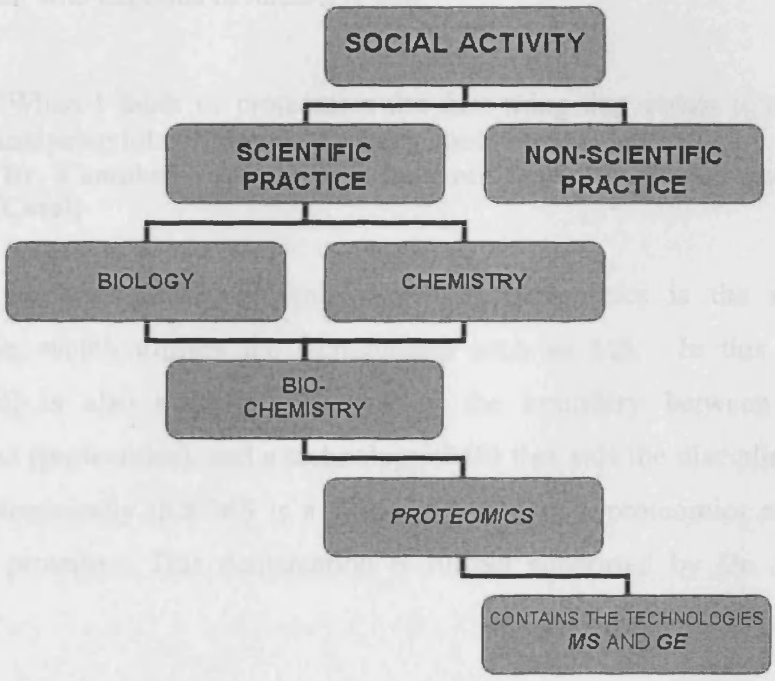

Figure 5.4: Interpretation of Liebler's Positioning of Proteomics.

In essence, Figure 5.4 shows science demarcated from non-science by a strict boundary. Reference must again be made here that it is not impossible to move from what is viewed as a non-scientific practice to an accepted scientific practice, but attempting to cross that boundary is difficult. For example Collins and Pinch (1979) have shown how mainstream scientists rejected parapsychology, a so-called pseudo science, as an orthodox legitimate scientific research area[62]. Contained within the border of mainstream science is another category called biology. The boundary of biology is itself delineated from other natural sciences such as chemistry and physics (refer to Figure 5.1). As explained earlier, however, other hybrid research areas have emerged and then stabilised in-between the more traditional disciplines, for example biochemistry (refer to Figure 5.2). If we are to follow Liebler's description of proteomics as a discipline, then a further boundary is formed which is emerging within the *new biology* but with strong *protein chemistry*[63] links (see Chapter Two). This discipline is called proteomics. What is more, the research area of proteomics contains the technologies of mass spectrometry and gel electrophoresis within its boundary.

This definition of proteomics is supported by a research respondent, Dr. Campbell, who explains in interview that:

> "When I think of proteomics the first thing that comes into my head is analysing lots of proteins by mass spec[trometry]..."
> **[Dr. Campbell: Doctor and Lecturer in Bioinformatics and Proteomics (Core)]**

Above we have Liebler's explanation that proteomics is the study of the proteome, which utilises the technologies such as MS. In this excerpt, Dr. Campbell is also making and marking the boundary between a scientific discipline (proteomics), and a technology (MS) that aids the discipline. He states quite categorically that MS is a visual aid used in a proteomics experiment to analyse proteins. This demarcation is further supported by Dr. Phillips who states:

---

[62] This is all suggesting that 'scientific' claims over knowledge are more legitimate than 'non-scientific' claims.
[63] Thus, explaining the positioning of proteomics within the boundary of biochemistry.

> "proteomics was a blossoming science..."
> **[Dr. Phillips: Senior Scientific Database Curator (Core)]**

Here, proteomics is explicitly talked of as a science as opposed to a scientific technology. In talking about proteomics as a science, it could be argued that Dr. Phillips also supports the proteomics discipline diagram.

Nonetheless, not all participants responded in this manner and the boundaries between (i) science and technology and (ii) tool and discipline become somewhat blurred when proteomics is claimed to be a buzz-word.

> "I would say proteomics...the distinction between that and protein research is a lot of proteomics is a buzz-word, and apart from doing protein research relatively high throughput, there is nothing intellectually novel about it as far as I am concerned."
> **[Dr. Edwards: Doctor and Lecturer in Molecular Cell Biology (Peripheral)]**

Dr. Edwards begins by claiming that proteomics is essentially a buzz-word. He states that he has experienced no change in the essential theoretical or epistemological basis from traditional protein research to proteomics. Instead, where he believes there has been progress has been in the development of new technologies, and specifically the emergence of high-throughput, automated machinery that can generate data on a mass scale. This paints a slightly different picture to my interpretation of Liebler's definition of proteomics, since Dr. Edwards suggests that there is a blurred boundary between proteomics and protein research. Dr. Edwards believes that '*intellectually*' proteomics and protein research are not distinct, and that the only significant difference between them has been a movement from lower-throughput technology to more high-throughput technology. Using Dr. Edwards' interpretation of proteomics there is an argument to be made that traditional protein research (emerging on the precipice of biology and chemistry) is the discipline or research area and that proteomics is just a name used to record a noteworthy development in technologies. If this is to be accepted Figure 5.5 could be used to illustrate their relationship.

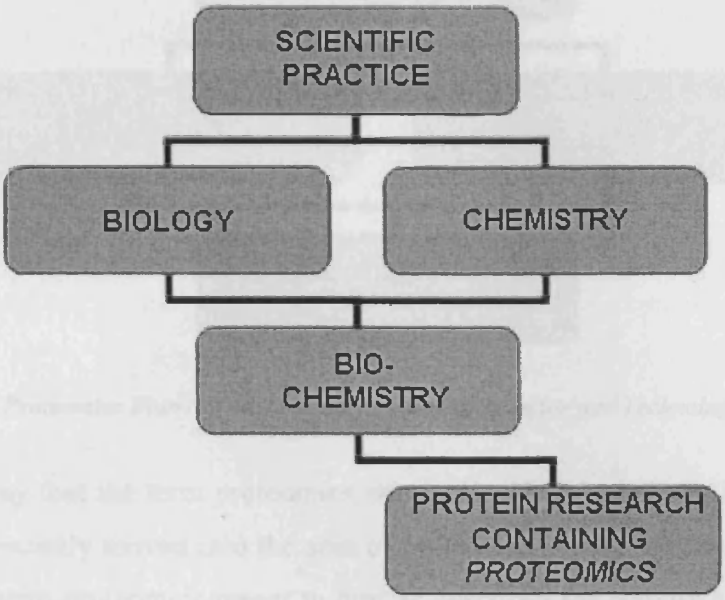

Figure 5.5: Interpretation of Dr. Edwards' Positioning of Proteomics.

Dr. Edwards' interpretation of what proteomics involves poses the interesting question of whether improvements in new technologies have been officially documented in the literature as a re-branding of the old term protein-chemistry with the new 'buzz-word' proteomics. This term would not encapsulate anything conceptually novel, but rather it records a period in history when there has been a significant development in technologies. If this is the case the progression of 'scientific practice' is meshed with the development of technologies, which in turn leads to the coining of a new scientific discipline, in this case proteomics. This last point implies that proteomics blurs the boundary between science and technology and that scientific research areas and technologies are deeply entangled. In this sense proteomics is malleable enough to be both the discipline and/or the tool used to aid the discipline. Using the definition put forward by Dr. Edwards, Figure 5.6 can also be displayed.

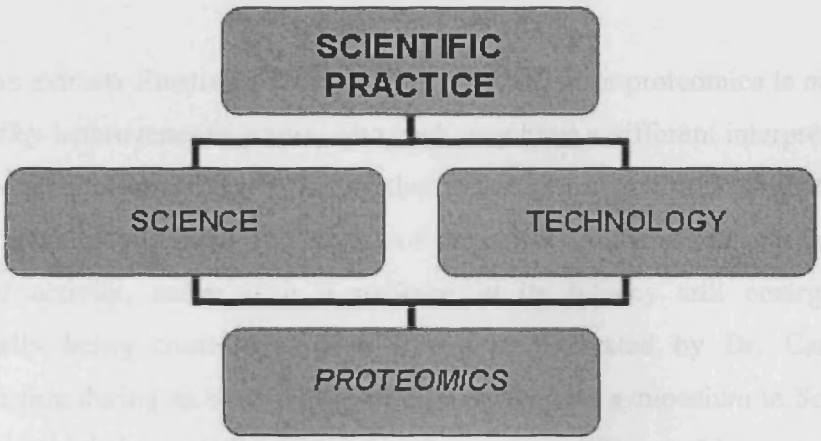

Figure 5.6: Proteomics Blurring the Boundary between Discipline and Technology.

A final way that the term proteomics was *malleably adopted* was by Dr. Morris, who has recently moved into the area of proteomics from statistics. When asked what the term proteomics meant to him he described the activity as being on the cusp of a dry laboratory science and a wet laboratory science:

> "I do look at a lot of applications and one of those is bioinformatics-proteomics. For me proteomics is about statistics and informatics, it is not all about biology."
> **[Dr. Morris: Professor in Statistics (Core)]**

As such rather than blurring the boundary between an intellectual activity and a technology, Figure 5.7 shows how, for some actors at least, the term proteomics obfuscates the boundary between computing science (dry lab) and biological science (web lab).

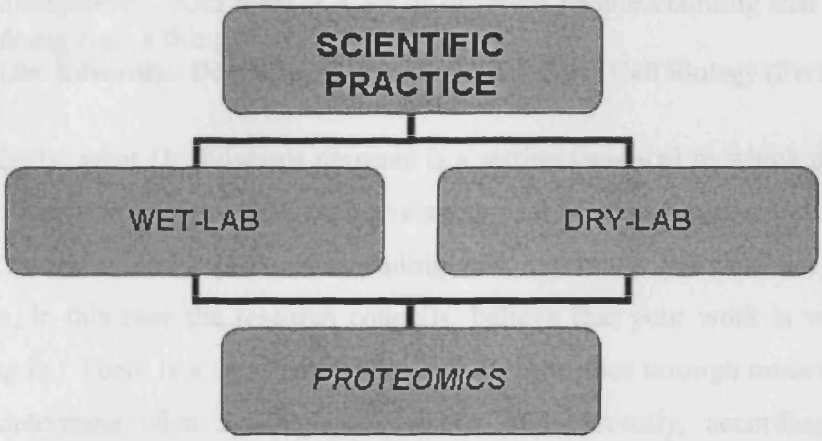

Figure 5.7: Proteomics Blurring the Boundary between Wet-lab and Dry-lab Science.

From the extracts illustrated it would seem that the term proteomics is *malleably adopted* by heterogeneous actors, who each may have a different interpretation of what the term means. That is to say that the term proteomics is seemingly in a state of liminality (Turner 1967). As of yet it has not stabilised into one clear, coherent activity, rather it is a package in its infancy still emerging and continually being contested. This is further illustrated by Dr. Campbell's proclamation during an opening speech of a proteomics symposium in September 2006, in which he stated that "proteomics means different things to different people" (*Fieldnotes* 2006). In response to this, a significant proportion of the audience nodded in agreement.

One of the reasons that it is malleable enough to be used interchangeably by different people as a discipline and then a technology is inherent in the notion that proteomics is a buzz-word. Dr. Edwards continues to talk about trends within science in general in the extract below:

> "Science is full of 'trendyness', as much as any other area of human life I guess. If you want to get funding and make people think your stuff is cool, you need to use the appropriate buzz-words of the minute, and they go in and out of fashion. And sticking omic at the end of anything, at the moment, is very common. It is not particularly intellectually rigorous, usually doesn't in the way that it is used, imply anything more than just doing lots of things quickly, and the extent to which all of these high throughput techniques really do give novel insights by being able to look at and compare across large datasets is largely, I would say at this stage, unexplored...And there is a lot of different people claiming that they are doing such a thing."
> **[Dr. Edwards: Doctor and Lecturer in Molecular Cell Biology (Peripheral)]**

Interestingly, what Dr. Edwards presents is a setting [science] in which the labels that you define your work have real, live agency. The classification and indexing of your work under a particular heading can determine whether the funding agencies, in this case the research councils, believe that your work is worthy of investing in. There is a clear belief here that science goes through transitory fads which determine what funding goes where, and currently, according to Dr. Edwards, omic science is in fashion. He also implies that to gain funding then the scientist must claim that their work comes under a particular vogue term or buzz-word, even when their work is empirically distinctive to the contingent written

definition of that word. The self make up of buzz-words and trends mean that what they entail must be fluid, changeable and heavily influenced by time. Trends come in and out of fashion and continually change, therefore the idea that proteomics is a constructed buzz-word may be an explanation as to why it is malleable enough to be adopted by different actors in different ways and at different times. Likewise, as Dr. Matthews recognises, actors define the term proteomics to fit the criteria of the funding agency:

> "Well there's lots of definitions [of proteomics] depending on which grant body you are sending your grant to...Proteomics to me is more of an in vivo look at proteins inside a cell...I take exception a bit to some of the structural people saying that what they're doing is proteomics if they're working on a protein in isolation...I'm not sure I see that as my definition of proteomics."
> **[Dr. Matthews: Doctor in Biochemistry and head of a Proteomics Facility (Core)]**

In this extract Dr. Matthews admits that proteomics has numerous definitions and that the way actors adopt the term is dependent on the funding agency that they are trying to impress. Despite the recognition of the term's fluidity, it seems evident that Dr. Matthews is uneasy with the ways in which some actors use it: 'I take exception a bit to some of the structural people". This may be one negative aspect of the term having such fluency. In the next section of the chapter, I critically evaluate the idea of the buzz-word suggesting it is the actor-defined term for what I call a proto-boundary object.

FROM BUZZ-WORD TO BOUNDARY OBJECT: THE PROMISE OF PROTEOMICS

Buzz-Word – "a keyword; a catchword or expression currently fashionable; a term used more to impress than to inform, especially a technical or jargon term" (American Heritage Dictionary of English Language 2007).

A buzz-word is a vague and a vogue word; it is often a neologism that is commonly used in technical surroundings. Its apparent flexibility could be compared to that of the malleability of an (objectified) boundary object. Specifically, in the rest of this chapter, I argue that a buzz-word is the actor-

defined term for what I call a proto-boundary object: an object that mobilises and attracts heterogeneous actors to its boundary within *phase zero* of a research area's development. Phase zero is the stage when initial murmurings of a scientific activity begin to get funded and organisational practices begin to gather momentum. Once the concept has been accepted by the majority of relevant actors as an area of scientific development, then the term might enter *phase one* of scientific development, where actions begin to speak louder than words, where the knowledge created begins to support original hype, and where a *proto-boundary object* may potentially begin to stabilise into a more robust *boundary object.*

Buzz words have the function of both impressing and obscuring meaning. Sometimes this obscurity is the result of its intentionally wide acceptance of various actors and interpretations, while at other times it is based on its own vulnerability of not knowing exactly what it is. The comparison to a boundary object can also be seen in its fluid, temporal nature. By self definition a buzz-word has to be temporal. It is a word or a statement that has *political* clout, but *political* clout often only for a limited niche period of time. This is supported by Dr. Edwards' earlier statement that "they go in and out of fashion" (p126). As time passes, however, original hype ('your stuff is cool' p126) that surrounds a buzz-word is often challenged by unanticipated problems and may be replaced by "varying levels of disillusionment" (Brown 2003, p6). Nevertheless, during the correct and opportune moment in time the word is able to gather momentum by attracting both new and established actors and technologies.

A buzz-word is also a re-branding of an existing more traditional thought or statement, and a concept that by its very nature creates expectation. The term can suggest promise and the potential of bigger, better and faster 'things' ("doing lots of things quickly" p126) since it is usually an important sounding phrase that is used to impress lay people by generally rejuvenating an existing or similar product[64], or launching a new one[65]. However, whether a term in science can live

[64] This supports the protein research to proteomics argument explained earlier.

up to its original hype is debateable (Brown 2003). This is an assertion supported by Dr. Cherry below and is a statement that the BBSRC/MRC are fully aware of, apparent in their 'hope not hype' stem cell research campaign (MRC 2008):

> "I think it always happens when something new comes out and a lot of money has been invested in genomics and proteomics and other omics. And there is a lot that can be derived from it, but whether it can live up to the wilder pieces of the hype is hard to say."
> **[Dr. Cherry: Bioscientist and Molecular Biologist (Peripheral)]**

This concern about hype and expectation in science is also expressed by Dr. Griffiths who acknowledges that there needs to be a stabilisation period within proteomics:

> "I mean it is not a massive problem but expectations run high and this has been a common problem throughout all the post-genomic sciences, indeed genomic science. You constantly hear in the media [that the] genome is going to solve the problems X Y and Z, [and in] ten years time we are going to have drugs for all these diseases, and of course, that is probably not true we are along way off these breakthroughs... Well I think my view is that perhaps, even more than other new technologies, I think that when you really start considering doing proteomics, we have a lot of interest from people who are reasonably naive to the area and the bottom line is this: it is still a very expensive technology, less so in terms of cost but more so in terms of the time that is needed to do the work."
> **[Dr. Griffiths: Reader in Bioinformatics (Core)]**

Dr. Griffiths admits that the word proteomics has raised anticipation and expectation of what protein research can achieve. There is a suggestion that this has been translated through the media as a breakthrough in drug targets, which in turn will have an effect on diagnostic research. This is the hype that goes hand-in-hand with a buzz-word. But Dr. Griffiths advises caution about this hype, claiming that the scientists who have been attracted to this *technology* are relatively naïve. In stating this, Dr. Griffiths, not only maintains that proteomics is a technology, but in using the term naive he is suggesting that scientists working in this area are rather inexperienced in using its techniques. This naivety could be explained as the result of the diverse actors that the term attracts, but it

[65] An example of re-branding in the UK could be the re-naming of the lottery. Originally called the 'National Lottery', it underwent a substantial rebranding in 2002 due to its dwindling numbers and has since been called 'Lotto'.

also represents the level of stabilisation of the research area. Interestingly, this was also an anxiety articulated in a major scientific journal. In 1999, Nature ran an article called *'The Promise of Proteomics'* questioning whether funding agencies should plough money into proteomics as a global activity in a similar way to what they did with the HGP. This question is followed by the sentence: "a boost now risks committing large sums to techniques that may soon be superseded" (p703). Thus it seems that buzz-words also bring with them an element of 'risk'. The risk being that the word is emerging and stabilising at a much faster pace than the actual activity. This worry is also apparent in an extract from Dr. Matthews:

> "The attitude of the masses to proteomics is not quite as glowing as it used to be. So there has been word on the street for some while that proteomics is yet to deliver; it has not fulfilled its early promise. The early promise was borne out of naivety that dealing with protein is nowhere near as straightforward as dealing with DNA, so where transcriptomics has given an awful lot of useful data, it's not that proteomics hasn't, it's just there is not that body of it. Therefore I think probably, the pendulum's swinging the other way [and] people are revisiting the biology....I think possibly what happened was that there were, certainly in the 2000 to 2003 era, a lot of technology developments...[But] translating it from that to actually being a functional tool that people can use was another step, which lagged behind more."
> **[Dr. Matthews: Doctor in Biochemistry and head of a Proteomics Facility (Core)]**

By implying that the hype attached to the research area almost got too far ahead of itself, Dr. Matthews' extract reveals a possible outcome of the fluidity associated with the term proteomics, and a reason why proteomics has not delivered its early promise. The result of this has meant that people are "revisiting the biology" involved in proteomics and are potentially proliferating the activity of proteomics into its past, present and future. Interestingly, Dr. Matthews once more talks about both the biology (the science) and the technological developments involved in proteomics activities, further suggesting that the term proteomics blurs the boundary between science and technology. Furthermore earlier in the interview she explains that it combines techniques used in a more physical science with biological understandings[66]:

[66] Once more this highlights the role of physical techniques in biological experiments.

> "It's absolute paramount importance that you can translate into biology [because] proteomics is a very technique based. It's quite a physical science; it's all about mass spectrometers and getting anonymous peptides to fly."
> **[Dr. Matthews: Doctor in Biochemistry and head of a Proteomics Facility (core)]**

It seems that, due to the vagueness of the terminology, a buzz-word's meaning may not match the conventional definition since it has the flexibility to take on its own form and structure. This definition deficit between meaning and form may widen if the attached expectation to a particular word or concept is not backed up by substantive evidence: "it's just there is not that body of it" (p145). In accordance with this dual entanglement of vagueness and value-expectation, the gulf between the 'rhetoric' and the 'substantive' (socio-material) aspects of the word proteomics could be rather wide. Early promises of what the term and all that it entails may achieve may not be viable, and may lay unfulfilled. Consequently buzz-words have the potential to sit awkwardly with some of the strict base principles to which science supposedly abides to. That is, funding agencies always have to fund scientific activities on some source of promise: promise that those applications that they receive will deliver. According to Lemaine *et al.* (1976): "in one way or another, all new areas of scientific investigation grow out of prior research or out of the extension of an established body of scientific and/or technical knowledge" (p2). But the leap from protein chemistry to proteomics or the leap from genomics to proteomics is arguably a larger one than normal scientific migration. This is because proteomics, as some authors have suggested, is a revolution in the production of knowledge (Chapter Three), and as such, any grant funding would be based on more than just an extension of previous work. It would also mean 'expanding' new methods, new technologies and new scientific beliefs.

Despite STS authors showing how the '*situatedness*' of scientific practices is often based on locally negotiated and locally manufactured knowledge and judgement (Collins 1992; Fleck 1979; Knorr-Cetina 1983; Latour and Woolgar 1979), previous Mertonian sociologists of science argued that science was founded on the principle of universalism. Universalism within science is the idea

that "the validity or value of any scientific statement is determined solely by the application of the technical norms of science; independent of the personal, social, political or national characteristics of the author" (Rothman 1972, p103). Thus grant proposals and the resulting distribution of grant money should be judged on similar universal criteria that the best proposal or the best scientists, independent of their group affiliation, should be granted the funding. However, buzz-words, especially those recognised and often supported by grant organisations (in this case twenty percent of BBSRC's funding is targeted areas) may get funding contributions based on the power and impact of a certain phrase and its embedded, but often miscalculated expectation. Thus, it could be argued, funding 'targeted projects' or buzz-words goes against this universal criteria, because in some instances rather than the value and quality of the work being the deciding factor as to what should be funded, it is conceivable that it is the *focus* of the work and the *re-branded terminology* which is assessed. Neutral value work (work not attached to a buzz-word) is then often judged unfairly because the heading under which their work relates to has not intentionally set out to impress the audience. This is a process which Dr. Harris also finds unfair when commenting on systems biology:

> "It is a policy statement as such. Somebody has all of a sudden decided that systems biology is important. I am not saying it isn't important but it is another classification by another name essentially. People have not been doing systems biology...Why is it all of a sudden now that it is crucial to have to study these things? It therefore takes resources away from other legitimate projects because it is earmarked as something that has to be done...It is a policy decision to fund this sort of research as a priority, as opposed to something that should be in legitimate competition with a lot of other work."
> **[Dr. Harris: Lecturer in Biochemistry Peripheral]**

Additionally, if the project or buzz-word fails to live up to the attached expectation, then funding becomes not only unsustainable but unjust and un-universal. The vagueness of the term means that what is actually being funded is equally as vague and poses the question of whether it could have been fairly judged to be funded in the first place. Thus, what I am arguing is that due to the notion of buzz-words, some social groups working in particular trendy sciences

do intentionally or unintentionally receive certain funding privileges, and as such, peripheral actors will align themselves with the trendy term.

RHETORICAL TACTICS TO GAIN FUNDING

The comparison of a buzz-word as a mobilising rhetorical force is seemingly supported by another proteomics actor who is a senior manager in charge of funding biotechnology in the UK. Dr. Harrison is fully aware of the rhetorical tactics used to gain financial support from different scientific actors who use the words in different ways to benefit their research and identity:

> "There is this continual impatience on the part of government, which is entirely understandable [because] the scales on which they think and on the timescales [at] which they move which is much, much faster than the timescales on which science moves produces certain behaviours...One of which is the continuous re-badging of what is essentially a seamless continuum. So the genomics brand name started to wear thin after the two spending review settlements. Yes that would be the best way of describing it. So post-genomics emerged and proteomics and various other things, which is essentially the ways of describing the same thing. But there are ways of conveying that the focus of the science and that the date of knowledge has moved on, and then you can explain that in different ways...
>
> You are right when you talk about it being labels. You have a phenomenon that results from this, [which] is that when things become buzz-words that have credibility with political decision makers, everyone jumps on the bandwagon and then the definition of them starts to become extremely broad and contentious. I mean systems biology is a very good example of that. When Gordon Brown (as Chancellor of the Exchequer) launched the ten year vision in the House of Commons he only mentioned one area of science in the whole talk, which I think was an hour long, and he mentioned BBSRC funding the centres of systems biology. And that was like a little red flag which meant that practically everybody who was wondering where they were going to get their next crust from suddenly discovered they were doing systems biology and it is as simple as that. I am not describing that in a cynical way, I think it is well understood by all the participants in this particular type of discourse that. That is what is happening...that you are using little phrases that encapsulate a very large amount of meaning as a kind of short hand between people..."
> **[Dr. Harrison: Senior Manager in charge of funding Biotechnology in UK (Peripheral)]**

Dr. Harrison's own language uses such expressions as 'buzz-words', 'labels', and 're-badging'. These terms stress the importance of the 'scientist's performance' in selling their work as a commodity. Once again, as his Gordon Brown[67] example illustrates, this is driven by the designated influential trendy terminology of the time and illustrates the real consequences that the term has. This was also emphasised at the launch of the proteomics symposium mentioned earlier. The first speaker, an academic doctor working in medical biochemistry and immunology, opened the session by stating that "proteomics is a buzz-word that has been around for a long time. The term can mean different things to different people". The fact that a single word has such a strong effect on the future of scientific research can show how and why different actors attempt to *malleably adopt* the term by either aligning themselves with the word or modifying the boundary of proteomics to fit with their requests and their self-identity. Dr. Harrison continues the discussion:

> "What does proteomics mean? Well to some people it is a tool, it is a research technology, it is a way of understanding all proteins expressed in a particular thing in a particular time, usually mostly mass spectrometric techniques. So that as itself is not an outcome. So you know, you might talk about proteomics in the same way as you talk about microscopy or something like that. Now you would get the odd person that would describe themselves as a microscopist or something like that, that probably means they are a technician and work on a big machine and people come along and use it. What we tended to do is use the term to describe research in which the technique is applied and that is a funny characteristic of the whole genomics area actually. In a way that is not the case that I have just described microscopy actually. You wouldn't find somebody referring to microscopy as meaning the actual things they are studying down the microscope and the research they were trying to achieve by actually doing it, but you would find people using it in that way in proteomics and I think that partly does come from your point about the label. You know it sounds like a good buzz-word. The point about it of course, is that if you actually...if you start to talk about proteomics research in this sense, meaning the understanding and knowledge coming out of understanding the proteome, in whatever context, then you are actually talking about a very integrated area of work that may involve proteomics research but also another number of things and also another number of other technologies like bioinformatics."
> **[Dr. Harrison: Senior Manager in charge of funding Biotechnology in UK (Peripheral)]**

[67] Current UK Prime Minister and ex Chancellor of the Exchequer.

Dr. Harrison further illustrates the malleability of the classification by describing the various ways that scientists have adopted, and indeed, adapted the term. He argues, and in doing so supports this chapter's analysis, that proteomics can be viewed as a tool, viewed as a research technology, viewed as the particular biology under study, and viewed as a dry lab activity (bioinformatics). This vagueness, he concedes, is borne out of the way that proteomics is used as a buzz-word, or as I have proposed, a proto-boundary object.

It would appear then that here we have a new buzz-term (proteomics) that when defined as a tool or technology by protein chemists or molecular biologists is malleable and flexible enough to be used by various different actors when they are in need of short term funding without them having to re-invent their already established identity. Unlike the argument of STS academics encroaching on the funding boundary of natural scientists, defined as a standardised package or a set of technologies rather than a discipline, proteomics can bridge these disciplinary boundaries more smoothly.

PROTO-BOUNDARY OBJECT AND PROTO-PROTEOMICS: THE IMPORTANCE OF FUNDING IN SCIENCE

It seems that the definition of proteomics is actor dependent. It is continually being constructed, re-constructed and co-constructed by heterogeneous actors. Proteomics constantly changes its shape and the silhouette that it forms is translated and manipulated in different ways by varying actors with their own agendas. For some, proteomics is used to define a discipline, for others a technology and for others still a set of technologies/tool or even an information science (Figure 5.8).

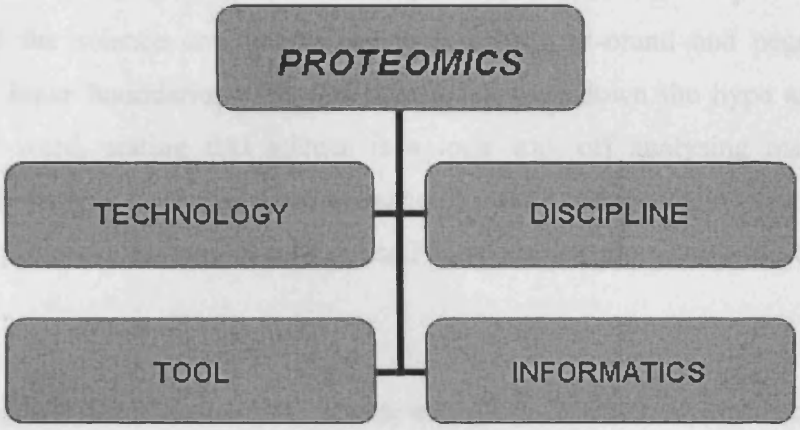

Figure 5.8: Proteomics' Different Identities.

This *malleable adoption* is also recognised by Dr. Campbell below, who despite stating earlier in the chapter that proteomics is the analysis of proteins by MS, a view consistent with Liebler's book, argues that actors have various interpretations of the term:

> "…I think throughout biology the same words mean different things to different people. Proteomics by definition are studies of proteins on a genome wide scale but actually the phrase these days is interchangeable with protein biochemistry. When I think of proteomics the first thing that comes into my head is analysing lots of proteins by mass spec, another person may equally think of bio-marker studies, another person may think of PP Interactions and another of protein chip sort of work. So it is a buzz-word but to be honest it has generally just replaced the two words protein biochemistry and equally people use the word proteomics to describe the study of individual proteins. Perhaps it is fair to say that proteomics is prote for proteins and omic for using technology rather than actually doing something on an omic level. They are using newer technologies and that is perhaps a better definition in some ways for the phrase because you know, in reality, we are some way off especially in mammalian systems to be able to look at things at an omic level."
> **[Dr. Campbell: Doctor and Lecturer in Bioinformatics and Proteomics (Core)]**

The extract is quite revealing and quite representative of earlier quotes. Dr. Campbell talks about proteomics as having both a singular and plural meaning implying its multiple uses, but also defines it in multiple ways. Proteomics is defined as a science ('biochemistry'), as a technology ('mass spec') and also as a paradigm shift in knowledge creation ('omic level'). In using words such as 'buzz', 'newer technologies' and 'interchangeable', his terminology is consistent

with those who see proteomics as a temporal fad within science promoting the idea that the science continually has to re-invent, re-brand and negotiate its dynamic inner boundaries. He also tries to dampen down the hype associated with the word, stating that science is a long way off analysing mammalian systems holistically. It is as a consequence of these heterogeneous interpretations of what proteomics actually entails that I have coined the term proto-boundary object.

I have argued, that proteomics *is* flexible enough to cross social worlds but as yet *is not* robust enough to keep one identity; a key criterion in Star and Griesemer's definition. Instead, respondents discussed proteomics as a discipline, as a temporal fad, as a technology, as a set of technologies, as an informational dry laboratory science, as a paradigm shift and as a re-branding exercise. I also argue that the level of robustness of the term is determined by the research area's level of development, and that robustness may come with scientific stabilisation since one lends itself to the other. The ambiguity surrounding the term 'proteomics' can also be captured in the term proto-proteomics. This term would reflect how proteomics has currently not stabilised to a level so as to be accepted by the majority of scientific actors involved in this type of work. For example it has already been illustrated that some actors view the term as just a 'buzz-word' and nothing really substantive. This revelation has also led me to suggest that proteomics currently exists in a *phase zero* of scientific development. Movement to *phase one* may mean proteomics becomes robust enough to have one recognised identity, and, in combination with the interdisciplinary nature of the field, will have the malleability to cross disciplinary and technological boundaries. If this is the case, on entering phase one of scientific development, proto-proteomics *could* stabilise from a proto-boundary object into proteomics the boundary object (the success story). Despite acknowledging that proteomics could stabilise into a boundary object, in proposing the concept of the proto-boundary object, I also inadvertently question two STS positions. The first is a statement by Susan Leigh Star herself.

In her paper, '*Power, technologies and the phenomenology of conventions: on being allergic to onions*', Star (1991) criticises Latour and Woolgar's Actor

Network Theory (ANT) for siding with the victor's position. She claims that: "they describe an order which is warlike, competitive, and biased toward the point of view of the victors" (p33). Her critique is that ANT does not take into account the loser position. In this chapter, however, I suggest that the concept of the 'boundary object' is also biased toward the point of view of the victor. In introducing to the literature the concept of the proto-boundary object I am proposing two possible trajectories. The first is that the proto-boundary object may stabilise into a fully developed boundary object (the victor's story), for example biochemistry. The second is the possibility that the proto-boundary object may remain as just a temporal rhetorical fad or may even regress into nothing[68] (the loser's story). Consequently I argue that the boundary object is also the story of the victor; a concept that highlights the success of order and which does not take into account the possibility of failure. Granted, Star is criticising a theoretical position, and I have criticised an analytical tool, however, the 'boundary object' concept has become so embedded in various forms of literature that is has become an accepted STS position.

The second (and perhaps more established) STS position I open up for further exploration is Collins' (1992) *'distance lends enchantment'* and Mackenzie's (1990) *'certainty trough'* theory. Both models describe how particular *ideas*, *truths*, and *facts* become to be widely accepted. Mackenzie's (1990) certainty trough supposes that uncertainty about a technological programme is greatest among the *producers* of the knowledge/technological programme, while lowest among those who are the *users* of the system. Consequently, he posits that the further away a group is from the production of the knowledge base the less uncertain (or more certain) they become. This is a similar concept to Collins' *distance lends enchantment* theory. Collins (1992) argues that as knowledge is funnelled outwards from the core-set to more peripheral groups the more certain actors become. Collins (1992) states: "distance lends enchantment: the more distant in social space or time is the locus of the creation of knowledge the more certain it is" (p145). It would seem, however, that when I asked both core and

[68] Some respondents have already claimed that proteomics over-sold itself in the late 1990s and early 2000s and, as such, failed to live up to its aspirations. As a consequence, the research area may now experience reduced funding, despite the current introduction of new technologies which may help it fulfil those aspirations.

peripheral scientists about their views on proteomics, despite all respondents being familiar with the term, it was not the case that the more peripheral actors were more certain of its virtues to create knowledge. Not only were some of the peripheral actors critical of how knowledge was being generated, but as several extracts have illustrated, they have questioned what the term proteomics actually entails. A few of the peripheral 'experts', for example Dr. Cherry and Dr. Edwards, are sceptical about what proteomics has achieved and what it will achieve in the future, and a few of the core researchers (Dr. Matthews for example) have expressed their concerns that the wider community believe they are not living up to their promise. Once more the fundamental reason behind this cloud of uncertainty is the *funding game*. Of course the core researchers have a vested interest in promoting proteomics, since it may help them achieve further funding and solidify their identity. On the other side, the peripheral actors may believe that the buzz-word of proteomics is overcrowding their opportunity to gain funding in their specialised fields by privileging other type of actors. This may be explained, in part, because no degree of closure (stabilisation) has occurred and that not enough time has passed; a caveat of Collins' position. Nevertheless, the argument that knowledge at a distance always feels more certain than knowledge just generated is questionable when you might have competing actors, competing claims on funding, competing theoretical positions and contested views on the development of science. Perhaps it is conceivable that those closest to knowledge production (the core) may be both the most certain and uncertain actors involved and that certainty may exist on either end of the scale (lay actors), while the 'peripheral expert' may always remain sceptical.

Despite more peripheral reductionist actors questioning the viability and future trajectory of proteomics, this chapter has also revealed that they are still attracted to the term. It is here we can use Callon's (1986) '*obligatory points of passage*' concept. There is a strong indication that if more reductionist-type scientists want to continue to be successful acquiring funding, they may have to acquiesce with the trendy terminologies of the time and the new methods proposed by novel omic scientists. Callon (1986) describes the story of the declining scallop industry in France to illustrate how knowledge is translated over boundaries. By focussing on two parties, three marine biologists and local fishermen in St Brieuc Bay,

Callon (1986) illustrates how the marine scientists attempt to entice and enrol the local fishermen into their ways of thinking and their methodological practices. Yet his analysis revealed how the local fishing group were not a homogenous one and he showed how the marine biologists failed to become an *obligatory point of passage*. In many ways the story of proteomics' emergence as a research area begins with funding agencies and research councils and their enrolment procedure as an obligatory point of passage. These organisations are critical network channels for scientists to continue practising science. Hence, once a buzz-word has been *accepted* as a targeted activity initiative (even if it does not fit with Merton's universalistic criteria), different scientific parties (for example omic and reductionist scientists) attempt to align their work to that term and in doing so they potentially acquiesce with the new methodological practice.

CONCLUSION: THE REAL CONSEQUENCES OF THE IMAGINED BOUNDARY

The chapter has attempted to position proteomics. In this regard it has discussed the idea that proteomics is a proto-boundary object, flexible enough to adapt to local needs but not robust enough to have one coherent identity. The chapter has also examined the relationship that scientific actors have with funding agencies which act as their obligatory points of passage. Viewed in this manner proteomics has the malleability not to pigeon-hole scientists. Unlike classification techniques and boundary demarcations that can construct identities (Bowker and Star 2000) if successful as a boundary object, proteomics has the potential to be able to continually re-invent itself so that scientists and technologies need not. This is because proteomics is able to bring already established heterogeneous identities on board with it. Functioning as a proto-boundary object, scientists and technologies do not necessarily need to change their identity but rather the object is fluid enough to be able to change its own identity to meet the scientists own needs. This is illustrated by the identity of actors that attended the September 2006 Proteomics Symposium. The second talk of the day was on Free Flow Electrophoresis (FFE), a technology used to reduce the complexity of protein samples. It was initially stated that the technology had been around for twenty to twenty-five years, and yet the emphasis of the talk was how it was now being

promoted as a 'novel proteomics tool' (*Fieldnotes* 2006). Essentially what is being described here is a twenty-five year old technology that is re-inventing itself and re-branding itself under the buzz-term of proteomics. What is interesting is that it is able to do this without having to change much of its own identity. Apart from making the technology compatible with other proteomic tools, FFE is able to permeate the constructed proteomics boundary quite easily and the transition from a supposed outsider into a welcomed insider is a smooth one. The chapter has also illustrated how the same is true for many heterogeneous actors[69].

But how are these boundaries really crossed? And, if proteomics crosses the boundaries of science and technology, the boundaries between paradigmatic shifts from reductionist to holistic science, and the imagined boundaries of constructed scientific sub-disciplines, then how does proteomics find its own identity? Without clearly defined and designated boundaries then the danger is that proteomics could become too lucid, too malleable and as such unidentifiable: it may never reach the status of a boundary object (the failure's story). As already stated, one of the functions of constructing imagined boundaries and specialised disciplines is to socially construct the role of an expert; legitimate actors who have the (scientific) authority to comment on and practise in specialised areas of research. But in an emerging and somewhat unidentifiable area of science how are those experts created and how do they manage to pull together the disparate identities of other scientists who are also permeating the boundary? These are questions that are answered in the four proceeding chapters. What this chapter has revealed though is that through socially negotiated and socially validated constructed boundaries the 'natural world' is ordered. These boundaries become real when they have real consequences on people. In the beginning of the chapter I stated that one of the functions of creating a disciplinary boundary is that it socially constructs an 'expert'. Thus it is then the boundary and not just the embodied skill that verifies whether one is an expert to a wider community. Even though the boundary is an imagined and socially constructed one, it has the real consequence of determining whether someone is an expert or not. In the case of

[69] Both omic and reductionist scientists have claimed to do proteomics.

proteomics the construction of a new boundary and a new field of expertise opens a niche for those who have an already established identity (often determined by their expertise) to gain a further and sometimes more prestigious identity by becoming recognised as an 'expert' in an area of research that is seen as cutting-edge or, as I have argued, trendy. As a buzz-word in biology, proteomics not only has the clout to mobilise different actors and technologies, but it can also endorse and advance their billing. The self defining feature of a buzz-word is that it attracts interest, and functioning as a proto-boundary object it also enables (for short periods of time at least) these interests to be realised. Yet how that field develops is, at least in part, determined by how the newly incorporated actors in the field build it, and that construction can only really be explained by the reasons why it is important to construct the discipline/field in the first place.

At the beginning of this chapter, I also explained why the constructions of boundaries were important in science for legitimising expertise and funding. This is no different for proteomics. Professional identity, credit and prestige are linked to disciplinary identity. For instance, it really matters to be recognised as being part of a discipline since it gives the scientists or the technology legitimation, professional acclaim and a sense of belonging. It matters in the sense that modern science is built on a foundation of finance, with funding being distributed to those people and parties who are identified as being part of a forward thinking (or trendy) research field. This all means that scientific identity really does matter in scientific practice and the construction of imagined boundaries have real consequences for both core researchers (the group situated within the boundary) and more peripheral actors (those often residing on the precipice). Therefore, the buzz-word could be seen as an instigator, or as a driving force in the development of biology. It is not a new idea or theory that always moves science forward in a cumulative way, since as some respondents have suggested, there may be nothing conceptually novel about proteomics. Viewed in this manner, proteomics is a re-branding and a re-energising of what existed before and along with the

development in high-throughput technologies is re-vitalising and re-forming the boundary around the new biology[70].

Thus it would seem that buzz-words have a large impact on the trajectory of modern science. But there is a further point to consider, and that is how the buzz-word is actualised. Here I would like to refer to Max Weber (1968) in which he described three types of social stratification. The third of these was the idea of a party. A party, according to Weber, is a group of people made up of an alliance of actors with the intention to struggle against other parties for resources. Modern science is set up in a way that almost demands that these types of parties exist[71]. In the post-genomic era, the story I have told here is one of the emergence of proteomics containing various Weberian-like parties who construct and re-construct its meaning in order to claim an identity. Groups of scientists malleably adopt the term in order to compete with other groups of scientists in attempting to claim funding as part of an obligatory point of passage. As a result of the advancement in biochemical science and the development of omic technologies, proteomics has been able to position itself in the geographical space situated between these scientific tensions. Furthermore, during its early stages of development (phase zero) and as a celebrated novel research terrain it is able to attract heterogeneous researchers from different locations and different boundaries. Despite this, however, none of the interview respondents defined themselves as proteomiticians or proteomiticists or other variants of that word. Instead, at the beginning of each interview, when I asked them to explain a little about who they are and what they had done, without fail they all defined themselves by more established identities, such as molecular biologist or a protein chemist. This further supports the idea that actors are able to keep their traditional identity while permeating the new proteomics boundary. Chapter Six will also address the reason why actors performing proteomics continue to align

[70] As I write this, I have just watched an advertisement, promoting the drink Southern Comfort. In the advert they are not promoting it as Southern Comfort, however, instead it is being publicised as SoCo. The drink may be in a different container and the name more catchy and quirky, but fundamentally Southern Comfort and SoCo is the same product. Yet it will be interesting to see if this re-branding has an impact on its consumers. Only time will tell of course, and the virtues will be socially negotiated among many, but already this re-branding has had an impact and began mobilising others, apparent in the fact that I am writing about it in this thesis.

[71] In modern science, different groups have to compete with one another to fund research projects.

themselves with their more traditional identities, and along with illustrating how standardisation is a stabilising practice helping to identify a research area, it shows how this is directed by a particular imagined future.

CHAPTER SIX:
SCRIPTING THE GOLD-STANDARD:
WHOSE STANDARD IS IT ANYWAY?

> *"Acceptance of prevailing standards often means we have no standards of our own."* [Toomer as cited in Bloom 1985, p3988].

INTRODUCTION

Standardisation and regulation are fundamental processes in the construction of legitimate, stabilised research areas. The creation of a *standard procedure* to perform tasks is often an attempt by an individual or a group of individuals to unify a community of regulated action, while the functions of the resulting standards are to routinise, manage and consolidate a research field. In this chapter I illustrate one story of standards generation by tracing the pathway by which the Proteomics Standards Initiative (PSI) constructs standards in *print* to be published in proteomics journals. In this respect, the standards act as a guide for identifying those researchers working in the field of proteomics. Furthermore, I assess how these standards in *print* are used as standards in *practice*. For instance, although regulating the research area is highly valued, a number of respondents in the study stated that the absence of a community standard has not prevented them from conducting proteomics work. Subsequently this chapter illustrates how standards in print, in this case standardised data reporting outputs, can be just as effective in aiding a research area to mature and stabilise as standards in practice, and that the printed standards script futures into current protocols.

Identifying what is involved in proteomics can be extremely difficult because as noted in the previous chapter, it would appear that proteomics does not have one single precise definition (Chapter Five). Furthermore, what actors *say* they do and what actors *actually* do are sometimes different. For example, some of the respondents in the study who clearly state that they *do* proteomics certainly *do not* perform proteomics in the way that some of the literature would define the

activity of proteomics. Equally, researchers who work in proteomics facilities or on proteomics projects do not label themselves as proteomiticians or proteomiticists, or any other equivalent term. From the thirty-one respondents who were interviewed in the project only one respondent used the word proteomics in their professional title, either calling himself a Proteomics Team Leader or Head of Proteomics Services. Consequently if the literature definition of proteomics is different from the empirically contingent definition of proteomics, what types of activity count as proteomics work? In answering this question this chapter illustrates the role that standardisation and standardising techniques play in the social construction and stabilisation of a scientific practice.

In the preceding chapter I examined whether the rhetorical function of proteomics as a buzz-word and a *proto-boundary object* (the notion of proto-proteomics) was dependent on the type of actors involved. I demonstrated that this, to some extent, *is* the case. One possible explanation for the *malleable adoption* of the term 'proteomics' is that the data reported in the preceding chapter is gathered from a mixture of both core and peripheral researchers conducting proteomics work. In contrast, the majority of data used in the current chapter is gathered from researchers whom I identify as being the core actors involved in cutting-edge proteomics work in the UK[72] and, in particular, concerned with standardising proteomics techniques.

More specifically, this chapter concentrates on the Proteomics Standards Initiative (PSI) whose aim is to create *data reporting standards* for proteomics outputs (introduced in Chapter Two). The primary function of the standardisation is to produce a platform and a format that may act as a common measure for enabling data comparison. The more general reasons for this type of standardisation are illustrated in the following data extract from Dr. Nielson, an actor on the periphery of proteomics work. He states:

> "There are two aspects [of standardisation]. From a research point of view it is important we are bringing this data together and it is comparable, and

[72] They could potentially be called the core-set (Collins 1992).

> from a diagnostics point of view it is not only comparable [but that] it is reproducible, reliable and accurate."
> **[Dr. Nielson: Reader in Molecular Haematology]**

Dr. Nielson holds that standardisation produces comparable, reliable and replicable data. These *scientific* virtues are explored in greater depth by Collins (1992) in his work studying the replication of the TEA-laser. In this chapter, however, I argue that there is also a secondary and indirect function of standardisation within proteomics: to produce an *ideal* for proteomics data reporting formats. By producing a homogeneous format that proteomics data reports should follow, the standardisation process ultimately helps in determining, legitimising and importantly identifying proteomics by creating a specific type of output that can be clearly recognised as a product of proteomics work. New intellectual fields need mechanisms of stability if they are to mature and blossom and therefore processes of stabilisation (such as standardisation) are key drivers in determining the future trajectory of a specialised research area. Using Akrich (1992), Eriksson and Webster (2008) also argue that standards are attempts to script futures into present protocols and practices. Hence, one potential outcome of stabilisation might be to transform a nebulous proto-boundary object (Chapter Five) into a more robust identifiable boundary object or sub-discipline. Consequently, the decisive message in this chapter is that the standardisation process aids in transforming proteomics from a fledgling proto-boundary object into a more prominent activity with an emergent prestigious identity and profile, but that this is directed by a particular imagined biological future.

WHY STUDY STANDARDS?

A major feature of Sociology of Scientific Knowledge (SSK) and Science and Technology Studies (STS) writing - whether it is an analysis of the disputes within and between scientists on divergent scientific controversies (Pickering 1984), the exploration of social processes of negotiation and compromise (Marrett 1987), or the landscaping of how boundaries are shaped and knowledge is ordered (Gieryn 1999) - has been research on how social consensus is achieved (or not achieved) within a community of actors. The definition, negotiation and validation stages involved in the construction of standards is another popular area

of interest for SSK and STS writers (Bowker and Star 2000; Eriksson and Webster 2008; Tanaka and Busch 2003), and another specific research interest in which actors are often found attempting to create a social consensus. It is interesting to contextualise this present study of proteomics within the history of these developments since this research is concerned with how scientists achieve a communal understanding of the world, and how this shared appreciation shapes the future practices of the research field.

One of the early, yet high-profile works on the explicit use of standards, and a study which introduced the notion of *Taylorism* into common parlance, was Frederick Taylor's (1911) *Principles of Scientific Management*, which examined the automation of the Ford motor company. In his study, standards are discussed as regimented procedures used in the automation of work protocols in order to increase product efficiency. His details of scientific management became so popular that *Taylorism* became the standard way to perform work in industrialised societies. This type of research on standards as a form of industrial and personal control dominated the literature on technology, work and protocols (Beynon 1975, Piore and Sabel 1984). However, with the emergence of science studies from the 1960s onwards, and developments in methodological techniques, such as laboratory studies and science ethnographies, studies on the explicit performance of standards have been conducted focusing on numerous other concerns. In many ways considering what the word *standard* represents, it is ironic that standards are utterly ubiquitous and have so many varied meanings, politics and trajectories behind their construction.

Among the most important of these explicit concerns studying the process of standardisation within an STS context include the work of Bowker and Star (2000), Clarke and Fujimura (1992), Star and Griesemer (1989) and Timmermans and Berg (2003). Their work has concentrated on, among other things, the consequences of classifications, the different types of standards that can be constructed, dominant technologies and products, and cross-boundary communication aids. David and Steinmueller (1994) have attempted to incorporate all the various meanings of the word *standard*, and as such, have indirectly illustrated all the differing foci shown in the examples above. David

and Steinmueller (1994) define a standard as "a set of technical specifications that can be adhered to by a producer, either tacitly, or in accord with some formal agreement, or in conformity with explicit regulatory authority" (p218). This definition unpacks the dictionary version of a standard: "something considered by an authority or by general consent as a basis of comparison or an approved model" (American Heritage Dictionary of English Language 2007), by recognising that standards need not be permanent or immutable and that the process of creating a standard requires a sense of shared agreement in a social environment.

If the 'standard' is what is measured or is the basis of comparison, the actual activity of constructing a standard is typically referred to as standardisation: "the process of making things of the same type all have some basic feature" (American Heritage Dictionary of English Language 2007). Timmermans and Berg (2003) define standardisation as "the process of rendering things uniform" (p24), and interestingly define the word *standard* as "both the means and the outcome of standardisation" (p24). To a degree, there has been less interest and less research by the STS community into the actual process of standardisation in comparison to the functions of the standard. I argue, however, that the emergence of a standard is a rich setting for STS authors, since the construction process of standardisation is a fluid, flexible, negotiable and contingent activity, whose social life is often determined by the numerous actors involved in constructing it. Furthermore, what it creates is often a static, boxed measure (a standard) that has or will have a direct impact on the future of research in the area (Akrich 1992). Consequently I aim to redress this balance by focussing on what standards *do* in the first part of the chapter, and then specifically concentrating on the *process* of standardisation in the second.

WHAT DO STANDARDS DO?

The chapter has already claimed that the creation of standards is a lived and embodied experience performed and shaped by potentially disparate actors, whose divergent views converge to create an objective bureaucratised document. This claim is consistent with the work of Stephens, Atkinson and Glasner (2008a)

in their account of the regulation of the UK Stem Cell Bank. Investigating the process of standardisation in this manner, it is possible to view standardisation as an inherently social activity that involves the bringing together of sometimes heterogeneous views, technologies, formats and communities into one accepted agreement. In this sense a type of universality is required to unite diverse subjective feelings into one objective culture. Timmermans and Berg (1997) maintain that achieving universality should be seen as a distributed activity and coin the term 'local universality' to address this transformation. The term emphasises how universality (the objective culture) emerges from localised practices (subjective accounts) and is "a product of contingent negotiation and pre-existing institutional and material relations" (p297). In this regard they argue that *all* objective universality must begin with some kind of *local* universality.

The roles of standards have been a source of considerable interest for Timmermans and Berg, and in later work they (2003) distinguish between different categories of standards and propose four *ideal type* standards:

(i) **design standards** – structural standards that ensure compatibility,
(ii) **terminological standards** – classification schemes that often develop into ontologies,
(iii) **performance standards** – measures created to achieve certain levels, and
(iv) **procedural standards** – protocols or organisational practices (pp24-25).

More specifically, design standards are intentional, specified techno-social components that are built into designs to create uniformity and 'mutual compatibility'. Terminological standards ensure that a concept or term is defined in an identical manner regardless of its cultural or temporal location. Performance standards are concerned with the outcomes, and in particular, what results should look like, and procedural standards set guidelines on how a particular routine should be carried out.

The standards that the PSI constructs involve all of the type of standards that Timmermans and Berg propose, although they are instantiated in different forms. For example, I have already argued that an indirect function of the standard is to

identify proteomics as a stabilising research area, this is because the PSI construct *performance* standards concerned with how a proteomics output should look. In addition to this though, the PSI also construct *terminological* and *design* standards since some of their work is about standardising proteomics ontologies, and in some cases the standards are digitalised and incorporated directly into technologies in order to regulate proteomics practice. Thus, Timmermans and Berg's model of standards helps to illustrate and evaluate the nature of PSI standards.

Notwithstanding the eminent virtues of their model, I maintain a further analytical tool is required to illustrate the function of standards. In Table 6.1, paying particular attention to the PSI, I argue that standards and standardisation potentially have nine key functions that can be divided into four sections. As we shall see, this model is compatible with Timmermans and Berg's (2003) earlier work, while adding further richness to the discursive tool available to us.

Using both the data collected here and past work on the role of standards I have identified four groups of standards and nine key functions (listed in Table 6.1) that could frame STS work on standards. I begin by describing bureaucratic standards, which I argue can improve logistical, compatibility and technical/organisational problems, and proceed to describe ordering standards that can help systematise and organise a research area, temporal standards that help to shape futures, and finally authentic standards that aid the regulation of a research field (Table 6.1).

Bureaucratic Standards
(i) **Comparative**: A standard can function by allowing two or more objects to be compared. It can do this by creating an intermediate object or by creating a measure that is compatible to all objects.
(ii) **Communicative**: A standard can aid communication and language variations in research areas by acting as a boundary object.
(iii) **Benchmark**: A standard can set a precedent that becomes the benchmark that is to be followed. An example of this would be a routine protocol.

Ordering Standards
(iv) **Uniting**: A standard can bring together and unite an area or areas of research.
(v) **Comprehending**: A standard can help create a sense of order and understanding in sometimes complex areas of work.

Temporal Standards
(vi) **Specific**: A standard can set a knowable and often realistic expectation to be achieved. By creating a measure that is universally agreed as the *correct* way to perform tasks it can provide actors with reachable aspirations; a set of guidelines or guidance.
(vii) **Directive**: A standard can shape the future direction of a technology or research. Once recognised as the dominant measure it can influence the future trajectory of that area.

Authentic Standards
(viii) **Ameliorating**: A standard can improve the quality of work or goods by setting a gold-standard or maximum level to be attained.
(ix) **Legitimating**: A standard can intentionally or unintentionally legitimise and validate a discipline or technology, while simultaneously filtering out others.

Table 6.1: The Functions of Standards in Scientific Work.

The four-part model of standards I have constructed is based on a different criterion to that of Timmermans and Berg who describe the different kinds of standards that exist: what the standard is about. Thus in their model, the first is about designs, the second about terminologies, the third performance and the fourth procedures. The model I produce, however, has a further analytical stage proposing nine categories of standards that define the standard's function: what they actually do in practice. Moreover, a particular standard may have more than one function. For instance it may be both *communicative* and *uniting* or both *ameliorating* and *benchmarking*. This, I maintain, could be because some of the functions are local while others are more global. As such I believe the framework I have developed enhances Timmermans and Berg's (2003) existing model since

it raises questions about not only the performance of standards, but also some of their actual purposes.

SCRIPTING IMAGINED BIOLOGICAL FUTURES

The types of standards I focus on in this chapter are what Timmermans and Berg label *design*, *performance* and *terminological* standards and their construction is intended to specifically perform functions one (*comparative*), two (*communicative*), four (*uniting*) and five (*comprehending*) from the bureaucratic and ordering categories. However, I suggest that the emergence of a collective standard in proteomics might also perform function nine (*legitimating*) of the authentic standards by intentionally or unintentionally sanctioning and identifying proteomics as a legitimate area of practice.

The aim of the PSI is to create community agreed standards for data reporting. The emphasis on community consensus is underlined by the PSI who state on their web-site: that their mission is to define "community based standards for data representation in proteomics to facilitate data comparison, exchange and verification" (HUPO-PSI 2007, p1). The explicit attempt to gain communal consensus through social negotiation and social validation supports the argument that standardisation within science is a social activity. Standardisation *must* entail an element of social construction when the process involves creating a measure within a community, and this argument becomes more apparent when organisations such as the PSI intentionally welcome as many people as possible from the community to contribute to its creation. This point is important because the word *science* is not a singular term, but rather a community-based activity that involves the collection of knowledges, theories, and technologies. Without this social interaction and social substantiation the word *science* could become superfluous and replaced by other words such as genuine or authentic. As such I trace the trajectory of the PSI standard by illustrating the number of *social* stages a PSI standard may travel through before it can become recognised as the dominant static standard in its domain (see Chapter Three). In order to do this I first need to locate where and how proteomics is situated in an imagined and visualised scientific/informatic model of biology. At this juncture I would like to

illustrate to the reader a particular past, present and future account of biology, which although not shared by everyone, is being used by modellers of science to direct towards a certain *imagined future*. This trajectory is imagined in three progressive stages or waves with current practices suggesting we are in stage two: the phase in which the standardisation of data reporting outputs is also situated.

The *past* account says that the world of biology was revolutionised with the mapping of the Human Genome (Welsh, Jirotka and Gavaghan 2006). Prior to this historic achievement, biology was a rather descriptive science conducted in smaller groups focussing on answering specific detailed questions. This reductionist approach to biology was made evident by the metaphor of the *gene* (Morange 2006). One of the outcomes of the Human Genome Project, however, was to demarcate biology into two terrains called reductionist and holistic biology, with an apparent increasing shift of attention from reductionist work to holistic work, and a switch in emphasis from the *gene* and towards the *genome*. This 'past' stage/wave is positioned as stage one and is called the *genomics stage*. The idea that this *imagined* account of science is progressive is supported by Faletra (2002) who argues that "reductionalist and fragmentary approaches, which typify a science in its childhood, are giving way to an era of synthesis" (p1), suggesting that the boundary of omic science is beginning to encroach upon the boundary of reductionism.

The *present story* of biology states that the world of biology has progressed from *genomics* into a second stage or wave called *post-genomics*. I suggest that this era is best characterised by focussing on communities, which while having individual coherences and stabilising practices, share traits that make them post-genomic. Improvements in technologies and developments in expertise and knowledge have meant there has been a dispersion of attention away from just the genome and towards other systems-based approaches as well. These additional biological systems include the transcriptome, the metabolome and the proteome. Morange (2006) states that the aim of post-genomic science is:

> "...to do more rapidly what was previously done in a very fastidious way. The objective, however, clearly remains the same as before - to explain

> the properties of the global system by precisely characterising its molecular components" (p358).

I have determined that the best way to imagine and order this stage is to focus on *scientific communities* and their social relations. The boundary of genomics has thus seemingly been broken and splintered to foster 'new' autonomous post-genomic communities mirroring the system that they are studying or the technologies that they are adopting. For example there are the metabolomics community, the microarray community, the transcriptomics community and the proteomics community. Interestingly, from originally including under its brand name all other omic sciences and all high-throughput technologies, today genomics can be seen to exist as *one* of these post-genomics sub-communities and as an area of research *within* the new larger dispersive post-genomic science stage/wave. Visualising the stage or wave in terms of *social communities* and *social relations*, renders it more accessible and comprehensible, since as Dr. Simmonds explains in the extract below, picturing the omic model of science is extremely sophisticated and complex:

> "The 'omics' things are horrible because proteomics and metabalomics sound distinct but they are not. About seventy percent of the techniques are common to both domains, which means if you're thinking about generating resources based on an omics view, you're going to do the wrong thing... There was a fairly clear dividing line between microarray, transcriptomics and proteomics. Since then they've got closer, because we've now got properly built protein arrays and antibody arrays, and things like that...Originally they were much more disparate so when you start suddenly thinking about metabolomics as well, 'you think hang on a minute we can't have mass spec[trometry] in twice just because there are two sorts of different people doing it'. The omic thing can be counter productive as we are beginning to merge together."
> **[Dr. Simmonds: Senior Software Engineer]**

Dr. Simmonds expresses the difficulty in trying to demarcate scientific disciplines from technologies, and suggests that new developments and subsequent implementation of technologies is actually bringing splintered disciplines closer together. He maintains that if you "generate resources on this omic view (the second model) then you are going to do the wrong thing" because some technologies permeate and traverse more than one boundary. Instead, Fujimura's (1996) *standardised package* concept might be a better way to visualise current

relationships, as her concept implies the clumping together of technologies, practices and actors. Furthermore, Dr. Simmonds' quote not only exemplifies the complexity involved in omic science modelling, but begins to introduce a third *imagined* stage or wave: a 'holy grail' era for informaticians where these demarcations are once again re-merged.

This final and desired *future* is a *systems biology* era visualised, in particular, by those involved in computational models. Within this imagined future, the aspiration is that the data derived, and importantly, prepared in all the post-genomic communities would be amalgamated and re-integrated to create a one-system approach to biology. In this scenario, *genomics* would have completed its full cycle and performed the re-badging of the seamless continuum from genomics to systems biology that Dr. Harrison explains in Chapter Five. If this stage is to materialise without problems however, Dr. Phillips argues that the data generated from all the different communities in phase two would need to be captured in compatible and commensurable formats so that if a systems biology approach becomes a plausible reality, then all the sub-communities generated data can also be coalesced smoothly:

> "The overall approach is that everything will be modelled in a functional whole from compatible data outputs."
> **[Dr. Phillips: Senior Scientific Database Curator]**

I argue that the potential futures of omic biology imagined by computer modellers such as Dr. Phillips are having *immediate* impacts on the identities of proteomics researchers, and if not yet effecting their current activities, will also impact upon their *future* practices. As such the need for standardisation within the proteomics community becomes two-fold. The approach I take is to group these needs into (i) a within-communities need to standardise, and (ii) a between-communities need to standardise. Both of these processes script futures into present practices. I explain the within-subjects need to standardise proteomics data first.

Within-Communities need to Standardise

The within-communities need to standardise is the concept that standardisation is required in order to vet, validate and unite the practices of the proteomics community. This is illustrated in the three quotes below:

> "It absolutely comes to a situation that you see which labs produce the data and depending on whether it's a well known lab or not, you either do or don't believe the data... I think the field is pretty much in a situation... where you can perhaps believe the data, or know the data from your own lab and possibly from your closest collaborators...I think the single most important thing that it's doing [the creation of a standard] is serving as a forum to develop benchmarks for what we perceive to be real data."
> **[Dr. Campbell: Doctor and Lecturer in Bioinformatics and Proteomics]**

In the first extract Dr. Campbell focuses on the need to unite the proteomics world by developing benchmarks that are clearly recognised as products of the proteomics community. This desire of creating a gold-standard or best-practice standard for the community is supported by further commentary from Dr. Matthews:

> "But you could imagine that there are two ways of looking at this. One is that [standardisation] is a good way of encouraging best practice. So if you are not terribly experienced, you can go back and look how other people set their parameters up [and it] might improve the datasets that you collect."
> **[Dr. Matthews: Doctor in Biochemistry and Head of a Proteomics Facility]**

While the goal is further developed by Dr. Phillips who explains how the PSI is trying to unite different groups within the proteomics community:

> "Four years ago there were no guidelines at all in as to what a good proteomics experiment was...there was a lot of variation in the metadata capture, ...sort of how the experiment was set up in the first place, a lot of variation in the detail where...they put the piece of software they use and...more importantly in a lot of ways the statistical analysis they subsequently did to do the identifications...Supplementary tables were fine at the time but they are starting to disappear in some journals up to three or four years old [and] the supplementary tables are not being saved any more...people were starting to make their own different databases, their own little websites and saving...the data online here, there and everywhere. Three years later the grant runs out, the group scatters round the country or all over the world to different labs and the IT department has had a clear out and the little database and the little website starts

> disappearing so there...has been a huge data loss. The whole remit of the PSI really is to do something about that so to produce standards so that one group could actually speak to another group and so the repositories could be built for the long term and the permanent storage of the data including the data at the time that may not have been deemed good enough for publication but on subsequent re-analysis three years later you may find that there is interesting information in there."
> **[Dr. Phillips: Senior Scientific Database Curator]**

The three quotes can be separated into the functions that the standards perform in relation to a within-communities need to standardise. To reiterate, I argue that a within-communities need to standardise is the requirement for a standard to emerge in order to standardise and regulate a particular individual research community. Furthermore, this standard has not been intentionally constructed to have an effect on any other research area/community. The first quote from Dr. Campbell illustrates how the PSI standards operate functions three (benchmarking), eight (amelioration) and nine (legitimation). The second from Dr. Matthews suggests that the standard executes function eight (amelioration), while Dr. Phillips believes it fulfils functions one (comparison), two (communication) and six (specification). What they all have in common, however, is that they rationalise standardisation as attempts to internally validate proteomics at stage two of the post-genomics model. By validating and authenticating data representation in proteomics, standards can set guidelines for what proteomics experiments' outputs should look like, and subsequently vet the mass disposition of data deposited in online genomic warehouses. The standardisation process can also internally distinguish proteomics as a legitimate and identifiable research activity.

Having dealt with the within-communities need to standardise, I now move on to the second concept I introduced above: the between-communities need to standardise.

Between-Communities need to Standardise

I argue that the between-communities need to standardise is directed by a *systems biology* imagined future. This is supported by Dr. Simmonds who explains how he is attempting to create comparable post-genomic data-reporting formats from

all post-genomic communities in order to be in a position to merge sets in the future:

> "So, just for the sake of argument, this could be proteomics, this could be metabolomics, this could be an array transcriptomics kind of thing, and in the middle there is the common bit. So this actually extends up to different degrees into different bits...and some will share a bit as well. This is a dumbing down of the proper picture, but broadly speaking what this is supposed to illustrate is that there is genuinely a common set of things. So words like 'experiment' can go in here, if you're thinking about the ontology, structures to capture project, design, inter-relationships between different collaborators in a multi-site project, or different technologies being deployed in the pursuit of one biological question, or whatever. So in terms of the format you've got stuff that goes in here as well. But also in terms of reporting requirements and stuff that goes in here. I mean if I get a fish out of a river then the fact that I can describe the river and I can describe the fish, I might want more description or less description depending on which particular domain I hail from, but there is, I would assert, a core set of descriptives that you'd find whoever was doing it. And then at the point, that you get your fish, you turn it into some sort of a paste and you start doing something with the gunge. Then you start to look at purifying the mRNA or purifying for proteins, or whatever, looking at small molecules, things like that. But saying that how and what the origin of the biological material was, is almost certainly common to an awful lot of different sorts of domains. So, again, in terms of reporting requirements there's a common bit there as well."
> **[Dr. Simmonds: Senior Software Engineer]**

In interview, Dr. Simmonds describes the modelling and informatics requirement standardisation. He simultaneously illustrated the process by producing a sketch which highlighted the complexity involved in this type of modelling. Essentially, Dr. Simmonds explains the usefulness of uniform standard formats across the different post-genomic communities. By using the fish and fish paste example, he suggests that different communities may want, or indeed require, more or less information depending on their background, but states that there will be a "core set of descriptives that you would find whoever was doing it". Dr. Simmonds is illustrating the desire for a set of common variables and categories across communities that are standardised in some sort of compatible and agreed data format. He further states that even though specific communities will each have slight variations in what they require, or how they display it, there is a base set common to all.

Argued in this manner, produced PSI standards would fulfil functions one (comparison) and two (communication) of the bureaucratic standards. The standards act as a type of boundary object, once again situated in post-genomics model two, but this time with a possibility it may lead to the stage/wave three model, *systems biology*[73]. In fact Dr. Simmonds is describing the circumstances that are required in order to move into stage three of the science/informatics model. His statement supports my argument that there is a between-communities requirement to standardise. If developments are such in technology and theory so that a whole systemic view of biology becomes a reality, then data from all communities have been prepared in a compatible manner in order to be amalgamated into a one system model approach of biology: the rebranding of genomics (model one) as systems biology (model three) via the current post-genomic (model two) stage.

FROM A SCIENCE MODEL TO A SOCIAL PATHWAY: A PARTICULAR FOCUS ON THE STANDARDISATION PROCESS

I have explained how the process of standardisation within the PSI is directed by a particular imagined *systems biology* future. I have also demonstrated how the process of standardisation is being performed as *normal science practice*, piecing together parts of the omic jigsaw within stage/wave two (*post-genomics*) of the omic view of biology. In the following section of the chapter I focus specifically on the *process* of standardisation.

Timmermans and Berg (1997) state that in the same way that:

> "things and humans alike follow trajectories flowing from their past towards possible futures. Protocols themselves have trajectories – they are constructed and reconstructed both by designers and in concrete use" (p276).

[73] Support for this argument can be found in a recent Nature Biotechnology (2007) paper by Taylor *et al.* exploring the range and coordination of a growing number of minimum information checklist standards.

After studying the processes of the PSI and analysing their types of actions I believe there is a level of chronology, marked by different phases, to the particular trajectory of the development of the PSI standard. Influenced by the work of Utterback (1996) on industrial innovation and Ravichandran and Sriram (2005) on proteomics, I produce an analyst's account of the construction process of the PSI standard. In this description I take the view that standardisation is a desired *stabilising* practice. This position is supported below by the manner in which Dr. Francis, when referring to the Protein Information Resource (PIR), nonchalantly talks about standardisation as an almost routine progression that all legitimate scientific research activities must pass through:

> "We had gone through standardisation in the early 1990s as a matter of course."
> **[Dr. Francis: Biotechnology Research Scientist]**

In spite of the statement by Dr. Francis implying that standardisation is a habitual activity, different standards in different arenas do emerge in (slightly) different ways. In the story of the PSI standard, I use illustrative responses from core proteomics interview respondents who indicate the standard has a potential twelve phase trajectory. The account is not meant to be instructional and I am not suggesting inevitability. Nevertheless, the standard's development course is directed by an *imagined future* and the phases outlined *may* reflect the development of standards in other scientific areas of research. Here I document the social pathway[74] of the PSI standard.

(Phase One) Identification/Juncture: My data shows evidence of an identification phase. This phase is when a person or a group of people identify a need for the creation of standards. In some instances this phase may occur as a result of a crisis within an existing research area, or in other instances it may occur at the beginning of the development of a new research area. Within the proteomics community it has occurred as a consequence of both. Nonetheless,

[74] I would like to emphasis that I used the term 'social pathway' since I maintain that there is a pathway to the standard's trajectory marked by some chronological order. Nevertheless, I am at pains to state that I am not defining this pathway as a model.

there is also an acknowledgement that proteomics technologies must have reached a certain level of maturity in order to address the issue.

In the extract below, Dr. Campbell argues that there has been no stringent standardisation of approaches within the proteomics community, and identifies a period of crisis in the development of the research area in which currently he does not trust the quality of data generated by other social groups:

> "There is a real concern that because there hasn't been any sort of rigorous sort of standardisation of the approaches that it's very hard to know the quality of that data."
> **[Dr. Campbell: Doctor and Lecturer in Bioinformatics and Proteomics]**

This view is shared by Dr. Phillips who believes that proteomics was suffering from the same complications facing any emerging discipline: the lack of uniformed communicative action:

> "[Proteomics] was suffering from the problem that all new sciences have, in that a lot of the basic work had been done very separately from different groups across the world, and they were having problems communicating at the data level, not necessarily at the personal level."
> **[Dr. Phillips: Senior Scientific Database Curator]**

The two extracts illustrate that the need for standardisation was identified at a specific period in time. The standard was required to aid communication (function two) between different proteomics groups in order to unite a research area (function four) by improving quality control (function eight) and setting benchmarks (function three). The aim of creating the standard was to identify proteomics as a recognisable and legitimate activity.

(Phase Two) Confirmation: My data also indicates the presence of a confirmation phase. This is when the identification of the need for a proteomics standard to emerge is endorsed by a group of people or an organisation. Below is an interview extract from Dr. Francis:

> "Once we had agreed on the principles of annotation and sometimes not everybody can agree on the principles of annotation and once it was

> conceded by the annotators in Geneva that yes we did need...standardisation, we could begin."
> **[Dr. Francis: Biotechnology Research Scientist]**

Dr. Francis uses the phrase "once it was conceded" demonstrating that the identification of the need to standardise was ratified or sanctioned by a wider group of people (the annotators) who acknowledged that standardisation was required. Once the identification for a standard is confirmed then individuals from the community need to be identified (and accepted) as the researchers who will construct the standard (*the standard creators*). It is at this point where I have identified a third phase.

(Phase Three) Reflection: Phase three is when several individuals from the proteomics community were either (i) identified or nominated as representatives on behalf of the community or (ii) have come together to form a stellar standards group in order to characterise what standards are required to be defined, and how they should go about constructing them. This is described by Dr. Simmonds:

> "You get a consensus from a group of scientists working in the field possibly through a society...anyone who is interested or they nominate people...and they hammer these things out."
> **[Dr. Simmonds: Senior Software Engineer]**

Dr. Simmonds explains how, in the reflection phase, actors interested in participating in the creation of the proteomics standard were given the opportunity and freedom to help construct it. He describes how together, they 'hammer out' what types of standards were required and how they proposed to proceed. This was the first of the deliberation-type phases in which members of the community articulate to other members their subjective feelings of what shape the standard should take.

(Phase Four) Financial: Once the initial agreement that a standard was required to embody the collective wisdom of a community was accepted (Lynch 2002), my data revealed a financial stage in which members of the stellar group attempted to attract funding. This is described in the first extract by a manager in charge of a funding council:

> "I got an e-mail one day, about two years ago, from Dr. Simmons who basically said I have just got this job. I invited him along, and he came to a meeting of the body I mentioned right at the beginning; the Cross Council Genomics Co-ordination Committee. And he just explained what he is doing. He was very good and he said I am...not here to ask you for money, I am just here to ask you if you could help me do my job. We invited him to speak at meetings we were holding of scientists, a couple of them, as a result of that...he put a grant application to us, which I believe was funded, although that was not funded by my committee, so I don't know."
> **[Dr. Harrison: Senior Manager in charge of funding Biotechnology in UK]**

The financial phase was required in order to gather further interest, mobilise actors and to illustrate to funding agencies that proteomics is a readily identifiable research field. In the interview above, Dr. Harrison describes the setting in which the scientists 'pitched' a successful funding bid. However, an alternative story is relayed by a member of the stellar group, one which suggests that their funding application was unsuccessful. When I asked Dr. Green to expand on his response that they had been funded poorly. He replied:

> "First of all the lack of funding, at least at the early stage, is not always a negative, for instance a well funded effort we get around to do standards, then this would have very much been 'oh the EBI is trying to push standards through'. While, as it is now the PSI basically is still a cohort hobby for everybody involved and then there is really a common interest to get this done because it is the best for everybody and the kind of internal agendas are really on a much lower level and so it is probably just as easy to come to a consensus if there is not just one organisation, which is well funded and tells everyone else what to do."
> **[Dr. Green: Proteomics Team Leader]**

The two proceeding quotes are contradictory. The first story suggested there *has* been funding for the project[75], while the second argues there *has not* been. In the second, Dr. Green even claims that the lack of perceived funding has not had a negative impact on the programme; instead, he believes it assisted the construction of a truly collaborative standard. Dr. Green argues that if a large amount of funding was given to the EBI organisation to produce a standard, it may have led others to perceive that they were enforcing the standard on the rest

[75] Although, Dr. Harrison did state that he only 'thought' there had been funding.

of the community. As it stands, however, and without this investment, he believes the standard will emerge more democratically as a truly community-based standard.

(Phase Five) Organisation: The fifth phase I identified in my data is the organisation phase. During this stage, frameworks and infrastructures were put in place whereby those involved in the reflection stage and other interested parties organise meetings and workshops to discuss ideas further. This is described by Dr. Phillips:

> "So the whole thing triggered off four years ago with two meetings here initially, although we have since moved all over the world, moving the meetings around. We were lucky, and I still don't know how on earth Dr. Green managed it [with] the other organisers. But they got into the same room data producers, data users, and most importantly manufacturers were involved right from the word go and were brought on board [as were] a lot of key experts in the various fields...As far as I know Dr. Green went around all the journals and all the people they knew and made up a list of all the people who should be there and invited everybody and just stood back to see who turned up... Every single meeting is completely open, anyone can turn up, [and] there is no charge. All they have to do is pay their airfare and their hotel but the meetings themselves are completely free."
> **[Dr. Phillips: Senior Scientific Database Curator]**

In the first part of the extract, Dr. Phillips describes what could be accepted as the reflection phase in which an initial meeting was organised. Later she portrays a setting in which subsequent meetings were organised by Dr. Green mobilising a plethora of heterogeneous experts from different research fields. I argue that the organisation of a myriad of interdisciplinary experts to contribute to the construction of the standard can clearly be identified as the organisation phase.

(Phase Six) Refinement: My data suggests a further consideration phase which I define as the *refinement* stage. This period is a fluid and continuous phase where the original concepts and ideas of the stellar group are discussed, deliberated and modified (possibly through e-mail or at conferences) by the wider proteomics community. This deliberation helps to create a more robust community-based standard. In the first extract Dr. Simmonds describes the story of how a

proteomics community member was keen for the PSI to be more meticulous in their characterisation of the definitions in official print. The group accepted this comment and as such frameworks began to take shape:

> "I mean he's basically an engineer so he's trained to do that kind of thing, rather than scientists that are trained to sit there and think: 'hmm'. Anyway, he has been involved with several different [standard definition] efforts in various domains and so he was quite eager that we tighten up our language a lot. So the MIAPE paper that will ultimately make it into print somewhere soon is much more rigorous in its definition of exactly what's meant."
> **[Dr. Simmonds: Senior Software Engineer]**

In the second quote Dr. Nielson describes an instance in which he attended a microarray consortium on the creation of the MIAME paper:

> "There are guidelines for microarrays and I have been on a microarray consortium where we have actually just come out recently with guidelines to be used in leukemia microarrays and what sort of quality of ions you should have and what sort of quality of cRNA [and] how you should store it and what the things are..."
> **[Dr. Nielson: Reader in Molecular Haematology]**

Both the above extracts are descriptions of how a standard and guideline becomes more robust and sturdy. As further actors comment on the framework, the standard may be moulded into an accepted boundary object (function one). In the first extract, Dr. Simmonds discusses the significance of a member of the community whom he states has different skills and expertise to him. The person he is referring to is an engineer, and therefore Dr. Simmonds believes, he is trained to scrutinise the precision of the definition. In the second extract, Dr. Nielson tells the comparative story of MIAME and in particular describes a consortium that he attended in which they discussed what quality of ions and cRNA they should use. Both of these extracts are describing the refining process of the standard by a further and wider cohort of relevant actors who may have different skills than those originally involved.

(Phase Seven) Production: The seventh phase revealed by my data I have defined as the *production* phase. At this point in the standard's trajectory, drafts

of the standards refined in the deliberation stage are produced and disseminated by the core stellar group as *working* or early *draft* papers of the standard:

> "And the latest PSI-MI version will be published this spring some time [or] early summer. Mass spec[trometry] as I said had much more legacy data and a lot more people protecting their area and a lot more variation in their data types, and so they had a far more uphill struggle, but worked incredibly well and incredibly quickly and the MZ data standard came out. Although they have not formally published it, it has already been adopted by a lot of manufacturers and had a lot of usage already."
> **[Dr. Phillips: Senior Scientific Database Curator]**

In the first data extract Dr. Phillips explains that the *first* PSI-MI version will be published in the Spring or Summer. Dr. Phillips' use of the word 'first', suggests that there will be further versions to come. Consequently I have interpreted that the first version is a type of draft or working version of what will eventually be a final-type version. Dr. Phillips continues:

> "Proteomics [the journal] has been very supportive. We have published meeting reports there. The first two came out in comparative and functional genomics, the first two meetings were in there and then after that proteomics asked us to produce one every time so people inevitably have actually followed it and a lot of people actually write in and ask for further detail, or mentioned it when I have come over and have followed what has happened through the meeting's reports and the websites."
> **[Dr. Phillips: Senior Scientific Database Curator]**

In the second extract Dr. Phillips elaborates further by explaining where these drafts and meeting reports are published. In this case the 'proteomics community standard' will be published in one of the relevant journals of the community called *Proteomics*. This description highlights the relationship between two separate groups; the *standard creators* (the PSI) who help construct the standard, and the *enforcement agencies* (the journals) that then disseminate the information to the rest of the community. I expand on this relationship later in the chapter.

(Phase Eight) Transitional Production: Another phase to emerge from my data is the transitional production stage. At this juncture, the construction process is

opened up once more to include additional *peripheral* actors[76]. Interestingly the production stage may occur after the transitional production stage in the creation of some of the documents. I illustrate this phase with two extracts. In the first extract Dr. Johnson describes one of the ways of contacting a standard creator[77]:

> "MGED have an email help line. You can email them with a request for a change. What you have to do is present a term that you want to include and a point of reference...I found them very very good."
> **[Dr. Johnson: Doctor and lecturer in Biosciences]**

During the interview Dr. Johnson explained how the model format only had scope for two categories: the male and female gender categories, but had no space for the category hermaphrodite. In response he e-mailed MGED highlighting this *error*. In the interview, Dr. Johnson explains how MGED altered the format to include the hermaphrodite category. This is an example of an add-on or alteration through what I define as *informal translation interaction*: modification to the standard by an anonymous user through informal channels. In contrast the second extract from Dr. Phillips is an example of a type of *formal translation interaction*:

> "So once a new term has been agreed, she actually writes it with the correct terms and gets all the cross references for it, and adds it and gives it its accession number. But we have a committee who vote them in or out depending on whether we feel it is an appropriate term, and make sure it is added to the correct place and then, generally during one of the workshop meetings, we will have an afternoon where the committee deal with new techniques where it is very obvious where in the hierarchy they fit in, and the main workshop if we want to re-write an entire grant or move something around or think we made a big mistake and want to redo something. Then we will discuss that with all the delegates, and make sure everyone is happy with the new way of doing things."
> **[Dr. Phillips: Senior Scientific Database Curator]**

Here, Dr. Phillips describes a setting in which a formal organised committee vote in or out a particular standard, often deciding on whether they had made an *error* in the original version or accepting that developments in the field mean that certain elements of the standard required significant updating.

---

[76] These may include peripheral actors commenting on and implementing further alterations and add-on features to the draft-standard document.
[77] Dr. Johnson was contacting the Microarray and Gene Expression Data society.

(Phase Nine) Substantiation: The ninth phase identified in my data is a stage in which the PSI collaborates with senior journal editors to substantiate the standard. This phase is described by Dr. Simmonds in response to a question about who they collaborate with:

> "In terms of academic and other sorts of collaborators, the journals. We try to stay in close touch with senior journal people when we're thinking about reporting guidelines. But this is in a more general way... Where we can find senior journal people who are also proficient in a particular area, we'll try to draw them in as a reviewer of these reporting guidelines that we make."
> **[Dr. Simmonds: Senior Software Engineer]**

And is supported in further commentary by Dr. Phillips:

> "We are now at the stage where things are being published, things are being put in place and the user acceptances have been pretty impressive so far [but] the people with the leverage are the journals and the grant holders."
> **[Dr. Phillips: Senior Scientific Database Curator]**

An official authentication of the standard has to be ratified by the relevant community's journals. As Dr. Phillips asserts: "the grant holders and the journals" have the "leverage" to implement the standards. This statement implies that although the standard has been created by an impromptu organisation within the proteomics community (*the standard creators*), this group *do not* have the power to encourage the rest of the community to adhere to it. Established scientific journals within the community (*the enforcement agencies*) do have the influence to enforce this standard though, not through coercion, but through specification. For example, if a proteomics-type article is to be accepted for publication it would have to follow the specific guidelines of the journal who may integrate a particular way of formatting into their publication requirements. In theory the proteomics actor does not have to abide by the standard formatting requirement of the journal, but in practice, they will do otherwise their paper or data will be rejected. Hence, the PSI works with the proteomics journals to substantiate the standard.

(Phase Ten) Stabilisation: The tenth phase to emerge from the study is the stabilisation stage. Here, after appearing in a number of different versions, the standard evolves into a mature, concrete and standard reference accepted by the community. Below Dr. Green describes how the PSI-MI standard has come to be accepted by *proteomics users*:

> "I suppose in the MGED conference in France two years ago now, I emphasised the very simple approach of the PSI-MI standards for molecular interactions and I had several people come to me and say yes that is great we want something simple and not overly complex and the next version will go down the same route. They will have less complexity, while we prefer the second version of the MI standard now, we are going for more complexity because we have established the lowest common denominator and now we need to extend to a somewhat more powerful standard."
> **[Dr. Green: Proteomics Team Leader]**

Here, Dr. Green describes a story in which there is a clear acceptance among the members of the wider proteomics community that he met at a conference that the standard created is practical and user-friendly. Once the users start to acknowledge its worth, it has the potential to stabilise into the standard reference. Dr. Green adds the caveat, however, that there is further requirement for a somewhat more powerful standard to act as the optimum reference. This serves to illustrate that stabilisation can be a very long process and involves both acceptance (for example by the users) and further negotiation (by the creators, enforcement agencies, and the users).

(Phase Eleven) Domination: I call the eleventh phase that emerged from my data and associated literature reading the domination period. At this stage of the standard's development, the standard becomes *black-boxed* and institutionalised as the legitimate, and sometimes only perceived way to act or perform in that area of research. A good example of this is the evolution of the QWERTY keyboard as an 'untouchable' standard in computing (Chapter Three). This creates a lock-out effect where alternative standards find it incredibly difficult to challenge the existing standard. Dr. Phillips elaborates on this universal standard:

> "Well if you can't compare data in the first place then you can't produce these reference sets so something had to be done. A universal standard needs to emerge in order to direct scientific action."
> **[Dr. Phillips: Senior Scientific Database Curator]**

Dr. Phillips justifies the emergence of an all encompassing dominant standard by suggesting that a standard is required to direct "future scientific action". Once more we can see how standards script futures in present actions. The extent of this rigidness, however, may be different between settings and communities. The PSI standard is a community-based standard and therefore there may be additional caveats to this penultimate stage in which the standard can go through periods of flexibility add-ons known as 'extensibles'. Examples of this are the transitional production stage and the below extract from Dr. Phillips:

> "So the idea is that the xml will stay fairly static but the controlled vocabularies will give you the flexibility with new techniques described by new controlled vocabulary terms which will still go under experimental method in the xml standard. So the controlled vocabulary terms give us our flexibility and our ability to stay completely up to date, but the xml schema will change when we want to bring a whole new concept in more than just a new method."
> **[Dr. Phillips: Senior Scientific Database Curator]**

By writing into the format amplitude for flexibility and extension, the standard is able to incorporate any contemporary concepts and/or shifting foci. Nonetheless, the core structure of the standard has been set and its foundations remain fundamentally the same[78].

(Phase Twelve) Reproduction: The final stage I identified from my data is defined as the reproduction stage. At this juncture, offshoots of the original standard have began to emerge focussing on very specific areas. Dr. Phillips explains how other groups are now active:

> "Other groups have opened up in the meantime, the gel electrophoresis group is now up and running and very active."
> **[Dr. Phillips: Senior Scientific Database Curator]**

[78] This supports the argument in Chapter Two in which an original standard template, such as the Linnaean taxonomic model, may remain the skeleton standard for years to come.

Once a community has created a standard, the framework that it is built on can be copied by other groups when creating their own community-based standard. An example of this phase is described in Chapter Three when MIAPE (from the proteomics community) based their format on MIAME (the microarray community). In the above example Dr. Phillips is explaining how the gel electrophoresis group is now established. At this stage, the standard could either be replicated to reproduce a standard in another community (as in the MIAME, MIAPE example) or it can be divided in order to reproduce new offspring.

It is not the purpose of this chapter to predict how the proteomics standard will develop. However to recap in tabular form, I have produced the following 'social' trajectory of the PSI proteomics standard (Table 6.2).

Stage Number	Stage Name
1	Identification/Juncture
2	Confirmation
3	Reflection
4	Financial
5	Organisation
6	Refinement
7	Production
8	Transitional Production
9	Substantiation
10	Stabilisation
11	Domination
12	Reproduction

Table 6.2: The Social Pathway of the Proteomics Standard.

I am not advocating a fixed, linear or chronological order from stage one to stage twelve in Table 6.2 since, as I have stated, there is an element of flexibility and fluidity in the standard's development. Instead, my portrayal of the standard's trajectory describes how the process of standardisation is inherently social and can be substantiated by numerous experts and groups. In everyday local practice there may be many loops, crossovers and even amalgamations of the stages

illustrated. Nevertheless, what I suggest is that for many community-based standards to emerge then the standard must travel chronologically through stages one, three, seven and ten. For example stage three should not occur before stage one and stage seven should not occur before stage three.

COMPETING STANDARDS

By unpacking the route and tracing the journey of the PSI standard to potentially arrive at the *domination stage*, I have shown how subjective, individual and embodied beliefs become disembodied and locked into the stabilised and objective culture of the social world. As a result of this type of institutionalisation the newly created standard becomes the bedrock by which the community can measure its activities. The standard not only aids the legitimisation of the research area by making it identifiable, but it also mobilises peripheral researchers with heterogeneous expertise and skills and attracts them to become part of its boundary. Hence, the story of how the standard is constructed is also a story of how core researchers attempt to mobilise peripheral researchers by creating a measure for them to follow. It is also here that we see the story of the disputes and the story of how scientists reach consensus.

Throughout my portrayal of the process of standardisation, I have described how an original gold-standard measure is socially constructed, and then illustrated how that standard is further shaped and re-shaped by other actors in the community, including the users, producers, agencies, communicators and translators. At the beginning of the chapter, however, I argued that much SSK and STS work has been focused on the formation of consensus when there are divergent perspectives. This is also the case in the life of the PSI standard. Conflicting expertise, different theoretical backgrounds and heterogeneous perspectives have to be negotiated when attempting to produce even something relatively as simple as an agreed term. This is illustrated by Dr. Fairbrother below who advocates the need for fluidity and flexibility when negotiating the naming of things:

> "But it is how you get different people thinking about things in that way...But you could take two very different research groups that will call the same gene by a completely different name just by the virtue of the

> discipline that they are working in. Now if you are a computer scientist coming in trying to standardise this: Who do you believe? Whose name do you choose? I sometimes wonder if we need to be less fixed of our understanding of that. Thinking of cars, generally everything out there is a Vauxhall, but if you see an Opal, well you know that is just a different name for a Vauxhall and you can cope with that. So why can't the system cope with 'well it is predominantly a Vauxhall but sometimes it is an Opal'; rather than going around and trying to find every Opal on the road and rename it as a Vauxhall."
> **[Dr. Fairbrother: Bioinformaticist]**

But in the specific case of the construction of the proteomics standard, negotiating divergent views was not just restricted to the naming of terminologies. A further story about competing standards between different parties in the proteomics community, and the separate roles and identities of *the standard creators* and *the enforcement agencies* also emerged from the data collected.

Interview respondents clearly recognised that the PSI was managing the 'proteomics standard', and yet the PSI did not have the community influence to encourage its uptake. Instead, the PSI were dependent on the community's journals for leverage and enforcement of its adoption. The desired relationship was that the PSI would create the standard by giving numerous actors in the community the opportunity to influence its construction. And once the community representatives were settled on its composition, the standard would then be substantiated when the journals integrated the standard as part of their publication guidelines. Nevertheless, despite a type of *social contract* being made between the PSI (*the standard creators*) and the journal 'Proteomics' (*the enforcement agency*) that the PSI standard would be used in the journal, the following extract describes how the journal published their own independent guidelines:

> **Interviewer:** So this incorporation of the proteomic journals, the users, specific guidelines of standardisation are of essential importance?
>
> **Dr. Green:** "That is of central importance but that is not what recently happened unfortunately. There is a competition amongst the journals also. And there is competition between MCP [Molecular Cellular Proteomics] who by-passed the existing community effort and put in this set of guidelines and now this hyper complex standard guidelines that have recently been published and basically they didn't take any notice of us or

> the detail the PSI gave and the same thing with the Proteomics [journal] standards, which have been published recently. They have something done where people would cite the Proteomics guidelines rather than the MCP guidelines. That is what nobody says, but to me that is very obvious. And in the same context I consider these much more useful because they are not overly complex but are more common sense guidelines."
> **[Dr. Green: Proteomics Team Leader]**

The situation described by Dr. Green is a further example of competing parties within scientific activities (Chapter Five). Ironically, in this case, the competing groups' intention was to create one single, unifying community-based standard, which it was hoped would unite a rather hybrid activity area of research (Chapter Five). Dr. Green explains that the PSI standard was not being published by 'Proteomics' since the PSI took too much time deliberating what the standard should look like. To use STS language, my interpretation is that the journey to stabilisation took too long and that the research area needed unifying more quickly. This view is supported by Dr. Green. In the extract below he is giving his opinion on why he thinks the PSI standard was not integrated:

> "Because we took forever for the overall guidelines...I suspect it was just time pressure and what the PSI does, which is to seek a very broad consensus, is extremely time consuming...I can understand it from the point that the journal has commercial interest and they have an interest in having the proteomics guidelines rather than the MCP guidelines and so they couldn't wait really so I suppose that is what happened."
> **[Dr. Green: Proteomics Team Leader]**

Dr. Green explains how he understood and accepted the journal's decision. This implies that even though the intention behind creating community-based standards is honourable, sometimes the need for a standard to emerge, by whatever means, may outweigh the need for it to emerge through a community consensus. To these actors, the lengthy twelve phase pathway of proteomics standardisation might be out-of-sync with the actual fast-paced nature of science, and that what is actually required is for the standard to progress in a quicker trajectory towards the dominant standard stage. If scientists want to make their area of research identifiable, it could be argued that a standard needs to begin emerging as soon as a desire for one has been identified. In this scenario a number of the phases would have been by-passed as the standard progressed from

stage one (the identification/juncture stage) to stage eleven (the dominant stage). Despite this, I still maintain that stages one, three, four, seven and ten are vital components in the creation of any standard. Alternatively, if the standard is to emerge democratically, consensually and universally I argue that it has to pass through more of the multiple phases. Moreover, in the long run, despite the initial time costs, a standard emerging through the social consensus trajectory may have additional benefits. A social consensus standard is, by its own definition, more likely to be accepted as a ubiquitous standard, while the dominant standard model may involve conflict or competition between two or more groups. Rather than unifying (function four) a community, standards emerging via the more *coercive*, instead of the *consensus* pathway, may eventually lead to a polarisation of the community. Any such division may mean that the standard would not fulfil many of the nine proposed functions described on p151.

CONCLUSION

Phillips and Pugh (2005) maintain that scientific method may be more usefully thought of as a way of documenting and writing research rather than as a way of actually doing it. They illustrate the difference between the academic articles produced by Crick and Watson (1953) in discovering the DNA molecule and Watson's (1968) book in which he described how it was actually done. Using the idea that today's biology is tomorrow's history, we might interpret Phillips and Pugh's approach to mean that what is written down and recorded is what is usually remembered and documented as science (Chapter Three). The (recorded) scientific method acts as a way of ordering any autonomous scientific activity. This emphasis on the importance of documents is consistent with Stephens, Atkinson and Glasner's (2006, 2008a) observation that twenty-first century cutting-edge science exists in a documentary culture. They make a distinction between what is 'doable' and what is 'documentable'; a similar idea to the difference between standards in print and standards in practice. At the beginning of this chapter, I stated that the absence of a standard had not prevented researchers performing proteomics. When asked if this had a major effect on his work, Dr. Campbell responded:

> "Not at present it doesn't have a major effect. I think that is very much from the point of view that in the areas that we are focussed on – [they] are fairly new areas. Whereas, it would be nice to sort of know which format to secure your data and to have it standardised, from an entirely selfish point of view it doesn't really make much difference. Because we are somewhere away from producing pretty large standardised repositories, in terms of us doing sort of doing our research, it doesn't really make that much difference. If there were those standardised repositories, sure life would be easier, but we don't so in terms of us being able to do our work, I don't know if it has such an impact to be perfectly honest."
> **[Dr. Campbell: Doctor and Lecturer in Bioinformatics and Proteomics]**

Despite the lack of a standard not impacting upon Dr. Campbell's present activities, in August 2007, the PSI published an article in '*Nature Biotechnology*' that described the method involved in developing a proteomics standard. The article entitled 'the Minimum Information About a Proteomics Experiment (MIAPE)', discussed the principles behind the need for adequate descriptions of proteomics experiments in relation to all other omic disciplines (Taylor *et al.* 2007). They argue that:

> "Reporting requirements for all technologies, protocols or entities that have relevance for many kinds of bioscience should therefore be developed in common between the relevant standards bodies (or by way of representative collaboration if no official standards body exists). In many cases, a 'tiered' solution should be sought (for example, for genomic sequencing, identify the source of the organism only; for proteomics or metabolomics, also give the feeding schedule; and so forth). To address all of these concerns, the PSI has become an active participant in the MIBBI project, which aims to anticipate or remedy such overlaps between sets of requirements" (pp888-889).

This statement frames the existing documentary culture in a particular way by prioritising the between-communities need to standardise. Taylor's approach also helps answer the question: Whose community's standard is being constructed anyway? Rather than merely being a proteomics standard it is clear that the PSI standard is a post-genomic standard. Thus in the same way that proteomics and bioinformatics may be viewed as boundary objects (objects that are able to cross boundaries and sometimes blur community identities) so too can documentary standards. Taking an omic view of biology, it is standardised data formats

(standards in print) that not only help to identify and potentially legitimise a research area, but effectively aid its stabilisation. This process of solidification is achieved by making inter post-genomic community formats compatible with one another (inter-boundary objects), and aligning the documented literature of proteomics with how it is actually performed. In this regard the PSI standard becomes a post-genomic standard, one that is directed by a systems biology future, and one where curators and computer modellers help to identify proteomics activities. Subsequently, if a lack of a standard does not have a major effect on, for example, Dr. Campbell's current activities, the creation of a standard does serve to *identify* him as an actor performing proteomics activities.

CHAPTER SEVEN:

COMPUTING BIOLOGICAL IDENTITIES

"Karma police, arrest this man, he talks in maths
He buzzes like a fridge, he's like a detuned radio" [Radiohead 1997].

INTRODUCTION

The mapping of the Human Genome marked a significant period in the development of a new type of biology. The movement towards what Liebler (2002) has termed the new omic biology has seen the reconceptualisation of biology as an informational science together with its traditional identity of a descriptive science. This evolution has given rise to the development of a number of what Fujimura (1996) may call *standardised packages*, and which I refer to in this thesis as interdisciplinary communities. The emergence and development of bioinformatics, a research area that combines the techniques and professional expertise of biology, computer science, computer modelling, mathematics and statistics is a key example of this shift. By focussing on the social implications of technologies, this chapter explores the development of bioinformatics as a freestanding identifiable discipline, and as a socially accepted division of labour in the post-genomic era.

Chapter One argued that biology is both a profession and an area of knowledge production. This is a similar distinction to Pickering's (1992) *science as practice* versus *science as knowledge*, which he uses to summarise the focus of early work of SSK scholars. More specifically, in his introduction to '*Science as Practice and Culture*', he argues that analysing "what scientists do is just as important as the knowledge they produce" (p7). This is also an interest for Penders, Horstman and Vos (2008) who evaluate the differences between what dry laboratory scientists *do* and what wet laboratory scientists *do*.

One signifier of the development of a research area is the maturation of its technologies. However, in some cases, knowledge inherent to a scientific

development or '*craft knowledge*' (Ravetz 1971) can become blackboxed within these technologies. Like all research areas bioinformatics is susceptible to this process of blackboxing. During the stabilisation stage, the distinction between biology as a *profession* and biology as a *knowledge producing research area* can become more apparent. It is also at this juncture that McNally's (2008) distinction between blackbox pessimists and blackbox optimists is useful. McNally (2008) identifies a blackbox optimist as one who argues the benefits of blackboxing knowledge within technologies. The multidisciplinary nature of proteomics is such that the diversity of skill required to master the field is daunting. Consequently, the construction of technologies such as bioinformatics may mean that certain scientific actors (for example biologists) would not necessarily have to learn new and time-consuming knowledges such as computer programming and statistical analysis since the technology would be able to translate data for them. On the other hand, blackbox pessimists believe it necessary to open existing blackboxes in order to evaluate how knowledge in such fields is produced and validated, since knowledge is often locked away in technologies[79]. This blackbox pessimist perspective is part of a larger argument and movement advocating the need for a more transparent scientific setting. In this chapter, however, I suggest a further potential blackbox pessimistic outcome. I argue that if bioinformatics is perceived as an area of knowledge production that can be captured within technologies, and distributed widely via the Internet as a form of blackboxing, then the status of the *bioinformatician* within the profession of biology is subtly changed. This is because the secondary knowledge producing technology[80] and profession can be transformed and reinforced as one which potentially could be performed by anyone. This has implications for distinguishing and rewarding professional positions within biology, and poses interesting questions about the role of expertise. Within this chapter, these questions are analysed by distinguishing between the role of the *bioinformaticist* (bioinformatics programme creators) and the *bioinformatician* (bioinformatics service providers), and by focussing on the knowledge gaps that exist in this emerging area.

---

[79] A good example is the Chang controversy, where Chang *et al.* had to retract five bioinformatics papers due to a faulty protein structure prediction (see Penders *et al.* 2008).
[80] With biology recognised as the primary knowledge producing area.

BIOINFORMATICS AND SYSTEMS BIOLOGY

In Chapter Six it was shown that present practices suggest that we are currently in wave two of an omic model of science. I termed this phase the *post-genomic* stage of science, which is consistent with other current writings on the subject (for example Diamond and Woodgate 2005). This wave involves many of the activities of Kuhnian *normal science* (1996), including the processes of sub-community standardisation that breaks boundaries, creates boundaries and extends localities. As the chapter latterly comments however, this wave is only the second tier of an *imagined* and *architectured* three-stage model; the final proposed *Holy Grail* phase in the model being called the *systems biology* stage. Interestingly, it is the phrase systems biology that seems to be not only the end stage of the model, but also the driving force behind the reconceptualisation of biology from its traditional role of a descriptive science towards its imagined future as an informational or digital science. For example, some may argue that the term systems biology has also influenced the development of a number of interdisciplinary technologies and communities in the *post-genomic* stage. These communities are what Zadeh (1965) might have called *fuzzy sets* or Fujimura (1996) called *standardised packages* since they have blurred but also maturing identities. They are blurred in the sense that technologies overlap more than one community, but maturing in the sense that the communities *have* attracted funding and mobilised actors. Examples of the communities given in Chapter Six include the microarray community, the transcriptomics community and the proteomics community.

In addition, the promise of systems biology has also seen the emergence of a technology that is fundamental to its progression. Ironically, the term used to describe the technology is sometimes used interchangeably with the term *systems biology* to mean the same thing. The technology is known as bioinformatics, the prefix *'bio'* derived from the word biology and *'informatics'* from the words information and science. Bioinformatics is a research area that combines the skills of biology, computer science and statistics (Chapter Two). It is also a

technology that symbolises the notion of the new biology as an informational and digital science.

According to Fujimura (2005), bioinformatics and specifically its meta term, systems biology "is the new buzz word, just as bioengineering was the buzzword in 1998 and genomics was the buzzword in 1990" (p197). Following this line of thought, I could have replicated a very similar analysis to Chapter Five (on proteomics) by concentrating on systems biology as a buzzword and fuzzy set, and bioinformatics as a boundary object, because the argument that Fujimura puts forward about the fuzziness of what systems biology entails is also supported by the extract from an interview with Dr. Harrison:

> "Okay, [the] BBSRC is very, very enthusiastic at the moment about systems biology, and we have a very clear view about what we mean by it, which is not the same thing as having a very clear view about what it is. If you look on our website, and find the calls for proposals for the two rounds of the systems biology centres, you will see that they avoid describing what systems biology is...I find it difficult to explain what systems biology is easily. One of the best examples I can think of is that if you think of a computer screen, if you wanted to try and understand or replicate or model the picture that was on the computer screen you wouldn't examine it pixel by pixel, reconstruct each one and then glue them together, which is essentially molecular biology. You would sample [or] you would examine a few pixels and then you would try and see if you could make the picture, and then you would look at what is bad about the model you made and you would go around and you would sample again. It is that process. It is a way of shortcutting the necessity to systematically test every single thing. In other words it is seen as a route to move from high-throughput experimental genomics activity, through to application by modelling and by the integration of data at different levels of biological organisation."
> **[Dr. Harrison: Senior Manager in charge of funding Biotechnology in UK]**

Dr. Harrison admits he finds it difficult to explain what the term systems biology actually entails and illustrates how the fuzziness in the term systems biology is very similar to the fuzziness in the term proteomics discussed in Chapter Five. The definitional difficulty is brought into focus by the fact that a National Institute of Health Bioinformatics Definition Committee (NIHBDC) had to be set up in 2000 to characterise the term. The committee incorporated nearly all definitions of the term in finally categorising bioinformatics as any "research,

development or application of computational tools and approaches for expanding the use of biological, medical, behavioural, or health data including those to acquire, store, organise, archive, analyse or visualise data" (Huerta *et al.* 2000). The establishment of this committee clearly indicates the complexity involved in defining and disciplining the new technological term. Rather than focussing on systems biology as a buzz word, or concentrating on bioinformatics as a boundary object however, in this chapter I explore the development of bioinformatics as a freestanding discipline and as a socially accepted division of labour in post-genomic science. I do this by critically examining the emergence of bioinformatics and the direction in which it is developing.

In order to discuss bioinformatics as a freestanding discipline it is first necessary to separate it conceptually from its identification with systems biology because as I mentioned above, the two terms have been used interchangeably. For systems biology I use Dr. Harrison's definition of the term outlined earlier: as a modelling aspiration of how biology could be conceptualised and performed. This also fits the three-part omic model's definition of systems biology discussed in Chapter Six. While for bioinformatics, I use the definition that Dr. Kennedy uses in an extract from his interview below: that is a kind of omic biology carried out on computers.

> "If you were defining bioinformatics, it is essentially IT for biological type data. And bioinformatics tends to cover genomic data, transcriptomic data [and] proteomic data."
> **[Dr. Kennedy: Bioinformatician and Cancer Informatician]**

Thus, in this chapter, I take the view that systems biology is the grander challenge of whole system modelling whereas bioinformatics is a smaller part of that large digital process.

LOCATING BIOINFORMATICS

Before beginning this analysis, I also position bioinformatics in the omic informatic model of science that I have discussed, and emphasise its dualistic relationship with proteomics.

In Chapter Six, I argued that the best way to visualise the *present* account of biology is to focus on communities. I provided the examples of the proteomics, metabolomics, genomics, transcriptomics and microarray communities to illustrate my point. I also suggested that the movement between waves was highly dependent on the development of new technologies. Bioinformatics is one of these key technologies, and had it not developed, there could quite conceivably be a completely different imagined model of omic science. The area has developed independently of the other scientific communities (Chapter Two), but in many ways still has a symbiotic relationship with them. Bioinformatics is the tool that stores and enables the analysis of vast amounts of omic data produced by the other communities. In this sense bioinformatics is essential to data generation, data storage and data analysis within all *post-genomic* communities (Penders *et al.* 2008). In interview, Dr. Campbell expands on this relationship by illustrating how proteomics is dependent on bioinformatics:

> "I think the bioinformatics in proteomics enables the automation of certain approaches. Even more for proteomics is genomics and the genome project and all of that information and that of course could not exist without bioinformatics. So I think often proteomics has moved forward without directly using bioinformatics. And that has only been because of the limitations in the amount of data we can get. But now that is increasing, it is getting to the stage now that absolutely it is much more reliant on bioinformatics. And certainly proteomics could not exist as it does today without bioinformatics. It would be a much lower throughput discipline."
> **[Dr. Campbell: Doctor and Lecturer in Bioinformatics and Proteomics]**

Interestingly, Dr. Campbell begins by declaring that bioinformatics is *in* proteomics as if it was part of the process of a proteomics activity. I will return to this point later in the chapter because it merits further discussion. But to continue, Dr. Campbell suggests that proteomics is approaching a stage in which it is becoming reliant on bioinformatics, especially if it is to be conceptualised and performed as a high throughput activity. As Bruun (2007) notes: "functional genomics, proteomics, metabolomics and many of the other new research platforms are based on the use of bioinformatics *tools* for storage, manipulation and analysis of data" (p187). Using the three-part model of omic science, the rise of bioinformatics can be interpreted as stage two of the model beginning to

mature and stabilise (*post-genomics*). This evidence supports the claim that there is a movement towards stage three of the omic model since bioinformatics may bring the communities closer together. It may also explain why some use the term *'bioinformatics'* and *'systems biology'* interchangeably, because aspects of bioinformatics are about analysing the information generated and producing it in a holistic form; a definition that could also be used for the term systems biology.

It is clear from the above statement by Dr. Campbell that bioinformatics is playing an increased role in omic science, and is beginning to mature into a fundamental research area in omic science. This is clearly one conclusion to be drawn from reading the literature and from analysing the modelling view of omic science. However, this chapter portrays a slightly different story and presents bioinformatics as a tool that while certainly being integrated further in omic work, is being integrated more as a *technology* rather than a *crafted research area.* Moreover, bioinformatics is often viewed simply as a service, providing biology with an analytical technique: a kind of specific *technology* in the division of labour of omic science. The consequence of blackboxing the knowledges involved in the research area means that if the term bioinformatics is indeed recognised and identified as being an integral *technology* in omic activities, and the driving force behind the movement to a *systems biology* future, then the role and the identity of the bioinformatician is not.

Crossing traditional boundaries and collaborating between disciplines is not a new phenomenon in biology (Fujimura 1996; Kay 2000). But, it is one of the key indicators of big science. For, as the science gets bigger it absorbs more and more disciplines, knowledges and identities. Bruun (2007) expands this argument in an article that identifies the challenges in bioinformatics. He describes how the discovery of the double helix structure of DNA by Crick and Watson (1953) was one of a number of success stories of how large-scale cross-disciplinary collaboration can lead to great scientific discoveries. Today, the new biology is subjected to a new cartography of epistemological coalitions that may include cross-boundary consanguinity, expert entanglement and technological treaties. The use of bioinformatics tools to handle mass data generation is an example of such a coalition. It is an activity on the intersection where technologies fuse with

sciences, where disciplines mix with services and where creativity coalesces with automation to produce *new* biological knowledges. Brown and Webster (2004) define it as a tool: "where massive databases are managed through high-capacity information infrastructures", and where there is a "convergence of the digital and the biological" (p2).

This chapter argues that the convergence of the digital and the biological must also include the convergence of digital and automated knowledge with biological and creative knowledge. Yet, as bioinformatics has begun to mature and develop, there has also been a polarisation of the discipline founded on notions of creativity. Consequently, the chapter explores whether bioinformatics is viewed as a service to biologists (a kind of machine), or whether it is viewed as a freestanding discipline where multi-skilled individuals with multiple knowledges attempt to bring order and clarity to a highly cluttered and complex area (a type of *demiurge*). Lash's (2002) work on 'science as knowledge' and 'science as information' is also pertinent here since the chapter will reveal beginnings of a divide within the field of bioinformatics predicated on notions of knowledge generation and information gathering.

DIVISION OF LABOUR AND THE DEMIURGE: THE CRAFTSMAN OR THE COMPUTER MACHINE.

The word demiurge comes from the latinised form of demioergōs meaning skilled worker. Literally meaning 'craftsman' it was used by Plato in Timaeus to describe the human creator of the world who fashions, shapes and moulds.

Big biology means an increased division of labour (Bartlett 2009). Contradicting Marx's ([1859]1999) '*A Contribution to the Critique of Political Economy*', this definition of labour must include not only work and labour but also the craft and tacit knowledge a *worker* has invested in the means of production. Correspondingly, big biology has also led to a greater division of knowledge, where knowledge has been increasingly shared between heterogeneous experts and different technologies at various stages of a project. This is essentially what Dr. Campbell means when he said that bioinformatics is *in* proteomics, it is a level or stage in a biological project where knowledge is transferred over to the

bioinformatician and their computer algorithms to analyse the data. But according to the bioinformaticians studied in this project, their role is often not being recognised and respected by biologists, and they are not identified as craftsmen who impart their creative knowledge in the process of analysis. Instead, they argue, they are perceived as a cog in the omic machine; an automated part of the process like any other routine high-throughput technology. This view is illustrated by Dr. Kennedy when he was asked whether he saw bioinformatics as a service or a discipline.

> "That is an excellent question, it is both. If you asked the bioinformaticists at my level, more often they will turn around and tell you it is a discipline. If you asked a biologist they should say it is a service. They probably won't, but they should because ninety odd percent of biologists I deal with or have...dealt with in the past, view, whether they know it or not, view bioinformatics as a service, and there is no doubt in my mind."
> **[Dr. Kennedy: Bioinformatician and Cancer Informatician]**

It is initially worth noting that Dr. Kennedy responded primarily with the gaze of a biologist and not a bioinformaticist. Fundamentally, his response is a protective retort that attempts to defend his area of expertise after attacking it by role-playing as a biologist. He states that bioinformatics *is* a discipline, but that biologists still view and treat it as a service. It is also important to note that he distinguishes between at least two different levels of bioinformaticians and that at his level (presumably the higher level which he coins the bioinformaticist), he views his research area as a discipline. This is something that I also return to later in this chapter.

When quizzed further on how he, as the bioinformaticist, perceives bioinformatics. He responded stating:

> "I view it as a discipline and it is a discipline because I suppose bioinformaticists at this level have their own research and are an interface between biology and computer science. So whether they may be biologists or computer scientists, which they tend to be, it is still at that interface. In that respect, it is distinct as a research area. So yes it is. I mean research needs papers and the papers that come out are bioinformatics in nature, and they can be applied. They can be computer

science in nature or they can be involved in novel biological type data. So it is cross-disciplinary in that respect."
[Dr. Kennedy: Bioinformatician and Cancer Informatician]

Dr. Kennedy strongly insists that bioinformatics *is* a discipline, since it has its own autonomous research area. This suggestion of *research* implies that the work conducted by bioinformaticists involves creative investigation, creative analysis and creative interpretation in the production of research papers. He also suggests that the area of research is highly interdisciplinary since it is also located at the boundary between computer science and biology and, as such, is an example of an emergent sub-discipline permeating the existing traditional boundaries of biology and computer informatics (Chapter Five). Nevertheless, it is particularly striking that in the third line of the extract, he identifies bioinformaticists[81] as either biologists or computer scientists as if they identify themselves with the more traditional categories, rather than those of the new sort. This comment is supported in an interview with one of the leading bioinformaticians in the UK. In this extract I have coded both the name of the person mentioned and their place of work:

Interviewer: "Perhaps I could start if you could give me a kind of background of what your title is and perhaps what you do? And who you are involved with?"

Dr. Griffiths: "So my name is [Dr. Griffiths], [and] I am a Reader in bioinformatics at the University [E], and I kind of head up our research group within the faculty. So our faculty is split up into sections and the sections into sub-sections and our sub-section is bioinformatics, functional genomics sub-section and I head that up."
[Dr. Griffiths: Reader in Bioinformatics]

Dr. Griffiths begins by stating that he is a Reader in bioinformatics and head of his sub-section. This was consistent with my original background research that he was a leading bioinformatician. Yet, two minutes into the interview, I realised that this was just his occupational definition in keeping with the department where he was employed, and was not how he would actually identify himself. The following extract is taken from the same interview:

[81] From here on I will refer to them as bioinformaticians, until I reach the section where I describe the distinction that is emerging between the two professions.

Interviewer: "Could you perhaps define bioinformatics?"

Dr. Griffiths: "As a scientific definition?"

Interviewer: "Yes or as your definition."

Dr. Griffiths: "I always struggle with this one. For me it is a biologically driven problem. Essentially I am a *biologist* that happens to use computers. I am more of an applied rather than a theoretical bioinformatician so in other words we tend to...we do develop our own software and things but I am not a *computer scientist* and I would never consider myself as one."

In response to being asked for a definition of bioinformatics, Dr. Griffiths states firmly he is *not* a computer scientist and would *never* consider himself as one. He identifies himself as a biologist who *happens* to use computers to do his work. Whether it is the case that Dr. Griffiths seeks to distance himself from being labelled a bioinformatician (and therefore as a serviceman), or the way in which people position bioinformatics, (as a sub-section inside a section inside a department) that forces Dr. Griffiths to authenticate himself with an established research area in order to justify his creative input (a biologist doing bioinformatics work), or whether he just lacks confidence in his bioinformatics skills, is unclear. Nevertheless, the fact that one of its leading researchers defines himself as a biologist perpetuates the labelling of bioinformatician as a less respected position within biology and as a sort of pseudo or proto research area.

In the extract, Dr. Griffiths also asserts that bioinformatics is a biologically-driven problem; thereby implying that it is not a computer-driven problem. This is revealing considering that Dr. Griffiths would not even define himself as a theoretical bioinformatician, but instead defines himself as someone who applies the tools (computers) of the trade to his work. Thus, despite being more of an applied bioinformatician, he believes strongly that bioinformatics is biological and not informatic in nature. This is consistent with the views of the other bioinformaticians interviewed. With one exception, all tended to have entered into bioinformatics from a biological background, and believed that it was the biological knowledge that was paramount to the research:

> "...but obviously coming from a biological background as well I can look at it more sensibly by not having to ask trivial questions constantly to people in the group because I have the biological awareness."
> **[Mr. Jenkins: Ph.D. student in Bioinformatics and Mathematical Biology]**

The division between computer skills and biological knowledge is one that also intensifies the polarisation of bioinformatics. As the interview extract from Dr. Fairbrother below reveals, very few researchers across the country have sufficient expertise in both the computer skills and the biological awareness to conceptually drive research in the area forward:

> "Let's say for instance we are supposedly solving these biological problems and all these sorts of things. But the computer scientists come in and it is more of the case of here is a problem that I can apply my pet technique that I have been working on in the past ten years. Oh! I can get some money to work on it but it doesn't have to actually have to produce anything useful....But it is quite difficult because I don't think the biological community has the expertise to drive things forward where it wants to be going. And if you can't drive yourself and you are being pulled by somebody who is not quite doing it for the right reasons you end up with this problem with what we are having. I don't think there is an obvious solution."
> **[Dr. Fairbrother: Lecturer in Bioinformatics]**

This is supported by further commentary from Dr. Kennedy:

> "There are very few people who have got the ability to think biologically and also to think of in terms of the computing needs for a or a number of projects."
> **[Dr. Kennedy: Bioinformatician and Cancer Informatician]**

And by Dr. Campbell who argues that most scientists are more skilled in one traditional research area than the other:

> "I don't think I have ever met people who have been equally interested in both and equally skilled in both. Almost by definition people do come from one discipline or the other. It is quite rare and unheard of, certainly in [the] top end of researchers, that they have been equally trained in both areas."
> **[Dr. Campbell: Doctor and Lecturer in Bioinformatics and Proteomics]**

And finally by Dr. Griffiths, who although paints a brighter picture of the future, still argues that on balance, bioinformaticians do not have an equal understanding of both fields.

> **Interviewer:** "Perhaps there are not that many people who are perhaps both au fait with the biological side and the computing side?"
>
> **Dr. Griffiths:** "I would say that it is becoming increasingly less true but it still is true on balance."
> **[Dr. Griffiths: Reader in Bioinformatics]**

These statements from four researchers working in the area of bioinformatics in the UK illustrate the difficulty of acquiring equal amounts of sufficient expertise in both the biological knowledge and computing knowledge required to drive research in this area onward. They all comment that very few, if any, bioinformaticians have the combined ability. This is inherently a problem of interdisciplinary research emanating from two or more traditional areas being merged into one sub-area. Each individual discipline has a long history of both theoretical and practical understanding. This means that attempting to be trained to an equal footing in both disciplinary areas is seen by many as simply not feasible, particularly in the early unfoldings of the research area. It also appears that both biologists and bioinformaticians consider training and qualifications to be of significant importance in this new emerging hybrid area (see Chapter Nine for more).

Nevertheless, despite the common consensus being that very few people are trained adequately in both areas, Dr. Griffiths believes that things are beginning to improve and that combined knowledge is beginning to increase, albeit at a moderate pace. This example is a clear illustration of the need to think about *knowledge* in addition to *work* when considering the increased division of labour within biology, since Dr. Griffiths admits to some sort of knowledge deficit within omic interdisciplinary biology. This knowledge deficit within the field of bioinformatics is the first of three knowledge deficits that I describe which are having an impact on the stabilisation of the research field. The second is to be found between the bioinformatician and the biologist.

Bioinformatics is a research area where knowledge is transferred over to the technology, and where biologists openly admit they do not fully understand the theory behind the algorithms. The fragmentation of roles, the abstraction of technology and the differentiation of languages and knowledges creates a knowledge gap in an omic biological project using bioinformatics. Weinberg's (1967) reflections on knowledge deficits in big science are useful here. He states that:

> "traditional working scientists are at the bottom rung – each one knows almost everything about nothing; as one progresses toward the top of the ladder, the subject matter becomes more abstract until one finally reaches the philosopher at the top who knows almost nothing about everything" (p47).

If we were to replace the term 'philosopher' with 'biologist' in this quotation, then once the data is handed over to bioinformaticians (researchers seen as lower in the hierarchy) the biologists running the project lose the intrinsic knowledge of how the data was generated. This leakage of knowledge is one of the reasons why the BBRSC has funded six co-located systems biology centres around the UK. Their hope is that the campuses will increase communication, understanding and knowledge-transfer between all of the experts and technologies involved in an omic biology project, and help prevent knowledge leakage.

When questioned about his understanding of bioinformatics, an area that is critical to his work, Dr. Nielson (a molecular biologist) describes the knowledge deficit that exists. The finding is also consistent with Bruun's (2007) position that "most bioscientists lack formal competence in bioinformatics, computer science, statistics and mathematics" (p190):

> "Yes I mean there are two questions there. Am I interested? Yes. Do I have the time? No. That is the big problem, I am an amateur bioinformatics person, [Dr. X] seems to be spending more time doing these things than me. I do want to know how these things are doing. I think it is important from the point of view of interpreting papers so you can understand the differences, the interpretations of what are happening and what genes are being expressed and how they selected these things out. So I think it is very important that you understand at least the minimum part of bioinformatics for any research so you can understand

> that. Do I have enough time and do I understand computing enough? Well, no to both of them. I understand a lot more but I don't have the time to practice on them. We buy in a couple of ready-made programmes like Gene Spring and we have got Array Assist and...I get ones downloaded...which comes from TIGR, the Institute for Genome Research, which is very good, and it is free which is even better. It is a matter of playing around with them and making sure you know how they work, but I mean how they go through and do all these things no [I don't understand]. The big one I don't know and what I should really know is R or Bioconducter from the R programme, and I feel as though somebody else should do that for me. But we don't really have that much bioinformatic support' really I think that is the lack in the university, for Gene Array stuff anyway, but that is a different matter."
> **[Dr. Nielson: Reader in Molecular Haematology]**

Dr. Nielson admits that he understands only parts of bioinformatics but he has no time, let alone the capability, to further his comprehension or to practice the skills. As a result, there is an additional third knowledge deficit; one between biologists and the blackboxed knowledge entangled within bioinformatics algorithms. A further example of the problem is portrayed in the extract below from an interview with Dr. Cherry:

> "...I will use bioinformatics as a computer. I will use bioinformatic programmes, but I don't know what the algorithms are and how they are constructed."
> **[Dr. Cherry: Bioscientist and Molecular Biologist]**

Dr. Cherry states that he does not know what the algorithms are, or how they are created. Instead he just uses bioinformatics as a blackboxed ubiquitous tool, comparing the way he uses it with how he would use a computer. These knowledge deficits, rather than highlighting the importance for the need of knowledgeable and multi-skilled bioinformaticists, actually reinforce the position of bioinformatics as a service, since biologists do not have the comprehension to appreciate the skills and knowledges used by bioinformaticians in their work. Consequently, rather than underlining how integral the skilled bioinformaticians are to developments in post-genomic science, the lack of expertise and the lack of understanding of bioinformatics work by biologists, creates and reproduces accepted divisions of labour between creator and service provider for scientists engaged in a uniquely twenty-first century technology. This is encapsulated in Dr. Nielson's comment on bioinformatics: "I feel as though somebody else should

do that for me...we really don't have that much bioinformatic support". Therefore, instead of being accredited as demiurges, bioinformaticians are placed in the same bracket as the technology; actors who others consider are automated pieces of technology designed to service biology, or to use Radiohead's lyrics at the beginning of the chapter "a man who talks in maths" and "buzzes like a fridge".

BIOINFORMATICISTS AND BIOINFORMATICIANS

According to the bioinformaticians interviewed in this study, the positioning of bioinformatics as a service is unjustified. They believe they are not only vital to the omic biology, but bring a skill set that is in short supply in many areas of the UK. During the stage at which I was recruiting participants for interviews in the study, I emailed a biochemist that had an interest in data analysis working in the school of medicine, about the possibility of an interview. Her reply supports the view that good bioinformaticians are fundamental to omic research. She wrote back with the response:

> "Yes this would be fine but not until the week after next! Will your work have an influence on the availability of bioinformatics around here?"
> **[Dr. Illingworth: Reader in Neuropsychiatric Genetics]**

During the interview she confirmed that she felt there was a lack of bioinformaticians in the University able to analyse the data being generated. When invited to expand, her response was couched in the rhetoric of bioinformatics being simply a service to biology:

> "Because I think some of the things I do, it would be much easier to have someone technical in post and to say to them, 'Please can you run this algorithm for me', or, 'Please can you code up this algorithm for me so I can try it out', or, 'Please can you pull this algorithm down from this website and see if it works', which I would find much more time consuming than they would. I think that it would be good to have more service bioinformatics. I think what would worry [the] college, perhaps they would not see these people being used as fully as they might be."
> **[Dr. Illingworth: Reader in Neuropsychiatric Genetics]**

The comment from Dr. Illingworth that: "I think it would be good to have more service bioinformatics" is an example of what Dr. Kennedy described earlier as

biologists perceiving bioinformatics as just a service, even when they recognise how fundamental bioinformatics is to the functioning of big biology, and to their own work practices. Dr. Kennedy firmly believes that biologists position all bioinformaticians under the category of a service, when in practice, as bioinformatics has developed, Dr. Kennedy and his colleagues maintain that a fracturing of the research area has occurred, into what they term *bioinformaticians* and *bioinformaticists*. Reference was made to this earlier when I remarked that there was a polarisation of the discipline founded on notions of creativity; this distinction can be identified in the following quotations. The first is from Dr. Kennedy:

> "There is this distinction that exists. I suppose it does, where a bioinformatician tends to be [on]the service side and a graduate MSc student,[or] maybe a Research Associate Fellow who has gone to work as part of a team, but does not have to come up with their own research. So essentially they are providing a service, a data analysis service...A bioinformaticist is viewed as, and these definitions are all mine, a bioinformaticist tends to be someone who actually carries out the research. So there is definitely a distinction. There are bioinformaticians out there and far more bioinformaticians than bioinformaticists, if that is how they are being termed. So yes that distinction does exist."
> **[Dr. Kennedy: Bioinformatician and Cancer Informatician]**

In the first extract Dr. Kennedy distinguishes between the research-orientated bioinformaticist and the service-based bioinformatician. This distinction is supported by Mr. Jenkins:

> "Ah yes I think bioinformatician, although there's...if you speak to my supervisor he would say there is a slight difference between a bioinformaticist and a bioinformatician. To be honest I am not really sure what the two are. I think of myself as more of a developer than a service, so I am not necessarily the person you would come to...it might appear that I am. I actually have done that initially in my Ph.D. and people come along and say I want to find this gene through an analysis of that. But it is an area I want to move away from and to move towards a developing aspect, developing applications rather than using them and giving people the results. I want them to do that rather than me. So that is why there is a definition between a bioinformaticist and a bioinformatician because they seem to vary."
> **[Mr. Jenkins: Ph.D. student in Bioinformatics and Mathematical Biology]**

And the same categories are recognised by Dr. Fairbrother:

> "People argue about the definition. I think the one that I am told I should use is bioinformaticist. The last definition I heard was that a bioinformatician is someone who uses bioinformatics tools, a bioinformaticist is somebody who develops them. So my background is that I started off in biology. I then picked up the computing as part of my Ph.D. and picked up the stats. So I have sort of now got a package of a bit of everything."
> **[Dr. Fairbrother: Lecturer in Bioinformatics]**

There was certainly a strong opinion among the respondents in the region of the UK in which I performed the research that a boundary exists demarcating two types of bioinformaticians. The interview extracts, all from bioinformaticians who position themselves at the higher level of the research, separate the term *'bioinformaticist'* from *'bioinformatician'*. They categorise the bioinformaticist as someone who uses their creative and technical knowledge to create new software packages that can analyse data. The role needs both *biological brains* and *computer competence* in order to develop biological software tools. This role is also in contrast to the bioinformatician who, according to their distinction, is someone that merely provides a service to biology by using already existing programmes designed by the bioinformaticists to analyse data for biologists. This type of categorisation is an example of what Price (1984) termed the *'role of instrumentality'* (p13), in which a new term is required to discuss the biological instruments that are created and utilised by researchers who need both technical expertise and craft knowledge to create, analyse and interpret any results that are generated from them. An extract from Price's (1984) discussion of this term follows:

> "We need a new term for these important techniques that help make new science. It will not do to call them instruments. Although the telescope fits this category, our term must let us include parts of the experimental repertoire that are labelled 'effects', such as the production of voltaic electricity, or the photo-electric effect, and such things as Cerenkov radiation or nuclear magnitude [sic] resonance. We must also include chemical processes, such as polymerisation and Lowry's method for protein determination and biological processes, such as recombinant DNA that lead to genetic engineering. I advocate the use of the term instrumentality to carry the general connotation of laboratory method for doing something to nature or data in hand" (p13).

In this sense, the technology provides the instruments for data exploration. Yet in the case of bioinformatics there also needs to be an understanding and manipulation of the technical element by human actors; first, in creating the technology (bioinformaticist), and second, in utilising and in interpreting the results (bioinformaticist or bioinformatician). Technical knowledge is required by trained professionals to provide the impetus to analyse and decode the information (Fine 2006). Even at the service end (the bioinformaticians), a technical approach is paramount. Dr. Illingworth's opening line in the interview extract on p194 admits as much: "it would be much easier to have someone technical in the post". Thus, although she portrays bioinformaticians as service providers, in effect, her use of the word 'technical' questions the boundary distinction that the higher-level bioinformaticists are making within bioinformatics. This is because the use of the word technical implies a sense of creative craft and tacit knowledge, even for those at the lower-level of bioinformatics (the supposed service providers). Dr. Harrison is also critical of the boundaries and the distinction of roles that are emerging within bioinformatics as researchers attempt to claim an identity. He comments:

> "When you have something very new that comes along, there is a sort of process of growth which is quite interesting and, I think it is probably true in proteomics too. When it first started, certainly when bioinformatics first started, people started saying this is all very interdisciplinary [and] we have got to have no boundaries, we are drawing on skills from all sorts of people, we are terribly eclectic and it is open to all comers and it is a very new field. It then starts to attract funding and then starts to develop a professional infrastructure of its own where people go to conferences and they meet one another and they start forming ideas of whether the people they are meeting are the same as them or not. You then reached a point where you suddenly had something awfully like the 'Amalgamated Union of Bioinformaticists' starting to say 'no we do it, you don't'. They put up little barriers and try to make sure they are fighting for their own corner, their own money and their own professional identity. The same thing happened to proteomics, but not quite to the same extent. The problem with bioinformatics is that it is not essentially biology so these were people in a field where they weren't doing the things that biologists did, they weren't doing experiments, not in the sense that biologists see experiments. So they were fighting the fact that they were up on the peer review panels so we would look at what they were doing and say this isn't biology, why are we paying for it? And other bizarre things like that."
> **[Dr. Harrison: Senior Manager in charge of funding Biotechnology in UK]**

Dr. Harrison believes that when the field began to emerge, heterogeneous actors were welcomed into the boundary of bioinformatics. However, since it has begun to stabilise and to attract significant amounts of money, separate factions have developed, which are fighting for their own identity and funding. Consequently, the opinion someone takes on the position of bioinformatics in twenty-first century biology is dependent on whether they view the biological end or the technological end, or a combination of both bio-technical knowledge as providing creative input into a project. Despite the roles of the developer (bioinformaticist) and the utiliser (bioinformatician) being clearly distinct, the future position of bioinformatics appears to be dependent on how biology views the wider category (without the distinction of roles) of bioinformatics. This is because biology is an established research area and the 'biologist' is an established research profession. Whether bioinformatics becomes automated or whether it will be viewed as a craft is crucial to how bioinformaticians claim an identity. This perception accounts for why Dr. Harrison believes bioinformaticians are putting up "little barriers". Nonetheless, the following example of a proteomics facility exemplifies that even those researchers and directors who are labelled as service providers are part of the creative process of how biological knowledge is created.

THE PROTEOMICS FACILITY

One of the interviews conducted during the research was with Dr. Strauss, a director of a newly-opened proteomics facility. The facility was designed to enable users to isolate and identify proteins of interest using state-of-the-art proteomics equipment. Although the strategic use of the word *facility* rather than *service* is telling, the *facility* does explicitly advertise the *services* they provide for biologists including, sample preparation, gel separation and mass spectrometry analysis, all of which they receive payment for. It could be argued that the development trajectories of both proteomics and bioinformatics have so far been very similar. One possible future trajectory of both would see the research areas becoming increasingly more automated. As the research areas mature and stabilise not only knowledge, but creativity becomes entangled, hidden and blackboxed in the technologies, and often the research areas are coined as services, facilities or techniques (Chapter Five). Although in the example of Dr.

Strauss' facility there seems to be an automation of proteomics analytical procedures and the facility is used by biologists as a service, in practice, as the extract below reveals, the director of the facility has a large creative input on most of the users' biological projects.

The first extract was in response to the question of whether she could describe a particular process of initial interaction between herself and her users:

> "A lot of people are trying to find differences between, say, a control situation and a treated situation, whether that's actual, you know, someone with a disease and I've got their serum, or whether it's cell cultures that have been treated different ways. And so what we tend to do, is first of all they'll come and talk to me about their project and we'll sit and discuss how we can address what they want to find out, and we'll sort of develop an experimental design together. I will sort of tell them the best way for them to make the sample for what we want to do. So then they go off and make their sample, bring it back to me and then I do all the proteomics."
> **[Dr. Strauss: Director of Proteomics Facility]**

Dr. Strauss states that she has a large input on the design of proteomics experiments, essentially giving the biological users *her* recommendation on the best way to go about their work. She developed this point further when asked about her role as service provider for the bioscientists:

> "I think it depends a lot on how much the people who are coming to me know about the field, because basically what they're coming to me for, I think, is my expertise in this particular area. So I'll have some people who come and they say 'oh yeah, we've done some 2D gels ourselves, etc., etc., but you've got all the machinery here'. And they don't really need my intellectual input, they just need someone to run the equipment for them. But then I'll have other people come and say, 'I don't know anything about proteomics, but I need to know what this protein is. How are we going to do that?' And then they're much more reliant on me to say 'Well, we could try this, this and this', you know, and 'I think it would be better if we went this route'...It's more intellectual input, but at the end of the day, the amount of, sort of hours of work, actual bench work would be very similar. So I think from the level of bench work I do, I personally feel I should go on people's papers."
> **[Dr. Strauss: Director of Proteomics Facility]**

If bioinformatics follows a similar trajectory to the proteomics facility example, there is the possibility that it too may be viewed as just a 'service'. Nevertheless, as Dr. Strauss' illustrates above, even at the facility or service level there is both a

technical and an *intellectual* contribution to the project, including an essential input into the decision-making process. It is at this juncture that the distinction between science as a working profession and science as an intellectual pursuit of knowledge is clearly visible. It is also apparent here that even if technologies do become more automated, there would still be a need for multidisciplinary researchers who have the relevant knowledge base, or as Dr. Strauss remarks, 'expertise', to make sense of the data generated in the area, and to direct users how best to conduct their research. When discussing the role that computers and automation play in biology Dr. Illingworth, who we have seen views bioinformatics as a service (p194), admits that most intellectual processing still goes on in actors' heads:

> "...the amount of processing they can do has grown exponentially over the last goodness knows how long; although it is nowhere near the level of the processing events that take place in most researchers' heads."
> **[Dr. Illingworth: Reader in Neuropsychiatric Genetics]**

Consequently, the 'gatherer', 'processor', 'provider' and 'interpreter' are fundamental if primary knowledge production is not to remain invisible within the research design. For, despite the level of bioinformatics processing increasing exponentially, it cannot compare to the processing that is apparent in the brains of scientific actors. It is for this reason that Price (1984) believes a new term must be created to describe this techno/social interface.

CONCLUSION

This chapter has argued that the importance of the *demiurge* cannot be underestimated. Decision-making, craft and tacit knowledge, biological awareness, technological advancement and the generation of knowledge is a dual partnership between human and machine in omic biology. Even at the lower levels of bioinformatics the term *service* does a dis*service* to the skilled researchers who have to cross a number of disciplines. If the role of bioinformatics becomes more automated, Price's (1984) discussion about the role of instrumentalities, rather than the role of instruments is borne out since technical and tacit competence will still be required to prevent knowledge deficits becoming knowledge blockages. Thus, this chapter suggests that the role of the

bioinformatician and bioinformaticist is more than a particular stage of labour. Rather, it is:

(i) the link between human and machine,
(ii) the bridge between descriptive science and informational science,
(iii) the difference between order and disorder and,
(iv) the connection in the omic science division of labour.

From this perspective, bioinformatics is the gateway that can unravel the omic labyrinth, and the bioinformatician is the gatekeeper that provides the maze map and the locksmith who cuts the keys in order to reveal the 'facts'. But to locate its current 'real' position in omic science, perhaps Dr. Campbell sums up the 'fuzziness' of the area best:

> "I think that you can only define it as a discipline if you are really doing some cutting-edge research and you are using entirely new statistical computational or mathematical approaches in the area that it hasn't been used before, and I think then it becomes a discipline. Otherwise I think it is a service or a facilitator for knowledge. Whatever the definition of bioinformatics, it depends on who you talk to. For me it's the use of computers to facilitate biomedical research, but other people will have quite different definitions so I think it could be everything. In University A I think there is a lot of people who do bioinformatics in terms of analysing data using fairly standardised approaches, there are very few people doing bioinformatics at the real research cutting-edge. For bioinformatics they are using bioinformatics generally as a tool or a combination of tools to look at the biological questions. So it depends on who you talk to. Very few bioinformaticists, I think, would take the view that actually it is a separate discipline. In a way how can it be a separate discipline because it is pulling from so many different areas? A nice way to think about it is a bridge between different disciplines rather than a separate discipline itself. But I wouldn't necessarily call it a service."
> **[Dr. Campbell: Doctor and Lecturer in Bioinformatics and Proteomics]**

In positioning bioinformatics as a facilitator of knowledge that helps to make sense of generated data, Dr. Campbell uses precisely the same word that is used to describe Dr. Strauss' proteomics centre, *facility*. Normalisation through stabilisation and standardisation make technologies such as bioinformatics more familiar and familiarity breeds conventionality. Conventionality, as Dr. Campbell

puts it, is not at the real cutting-edge of research, instead, it is just routine, fairly standardised Kuhnian normal science. In contrast, as a new technology or research area begins to emerge the creative role the actors play is usually recognised since the knowledge is unfamiliar. However, once that research field, and the technology stabilises, the creative knowledge becomes blackboxed inside the technology and the area of research becomes recognised more as a technique or a tool.

Subsequently, what this chapter has illustrated, is the role that the *'demiurges'* play in relation to the computer programmes that they create. Even when the technology has stabilised and matured, at the very least bioinformaticians facilitate the generation of omic knowledge which assists in extracting order from disorder. While, at best they are in the vanguard of cutting-edge biology, creating biological knowledge from computer data. Referring back to McNally's (2008) distinction between blackbox optimists and blackbox pessimists, the future trajectory of bioinformatics may be determined by the type of knowledge that bioinformaticians produce. The blackbox optimist may argue that boxing knowledge into a computer programme and thereby transforming it into a type of ubiquitous knowledge is a key indicator of a technology stabilising. What I have put forward in this chapter, however, is that the key term in the blackbox optimist's argument is the adoption of the term *technology*.

The bioinformaticians who contributed to this study believe that a considerable amount of bioinformatics is a research area or discipline, and that this must continue to be the case even when the technology stabilises. Moreover, if the research aspect is to gain better recognition, biologists must acknowledge the creative, analytical, and sometimes tacit knowledge input bioinformaticians bring to their work. If this is identified, then bioinformatics may emerge as a hybrid 'discipline' rather than a hybrid 'technology', which in turn will have a positive impact on the identity and professional roles of bioinformaticians. Nonetheless, the way that bioinformaticians in this study are attempting to claim an identity in the post-genomic era is to polarise their own discipline into the categories of bioinformaticists and bioinformaticians. Despite these categories distinguishing between the different roles, other biological actors, such as Dr. Harrison, believe

it is not achieving what they hope it would. Thus, rather than highlighting the knowledge gaps that exist within biological informatics, and emphasising the skilled nature of bioinformaticists, it is merely demonstrating that they are not performing experiments in the way biologists perform experiments. In turn, this may have a detrimental effect on their identity and how they may be funded in the future.

CHAPTER EIGHT:

MATCHMAKERS AND SPEED-DATERS: CROSS-COLLABORATIVE ARRANGEMENTS IN BIOINFORMATICS AND PROTEOMICS

PREFACE

Chapter Five outlined how an increase in interdisciplinary research within post-genomic biology has seen the more rigidly constructed boundaries of expertise that have been shaped over time being broken down, their contents shared and new boundaries drawn. Within the new biology it is hoped that the permeation of these traditional disciplinary boundaries can help move the current knowledge base forward. Using this definition it illustrated how proteomics and bioinformatics are interdisciplinary fields that have the malleability to attract and welcome a myriad of experts across various boundaries. Chapter Eight builds on this notion of the proto-boundary object and suggests that the new biology is becoming ever more interdisciplinary and a new type of researcher is required to navigate within the new domain.

During this research I have interviewed, read or been told about people performing proteomics whose professional expertise lay in biology, chemistry, informatics, computing, maths, engineering, and physics. The 2006 proteomics symposium referred to in Chapter Five was the second symposium of its type at that institution. The first symposium was attended by approximately seventy people, and according to Dr. Campbell, the composition was seventy percent from the medical school and thirty percent from biosciences. The second (which I attended) saw an increased diversity of affiliation, despite the number of attendees having dropped slightly, which interestingly supports the view that proteomics may not successfully stabilise into a boundary object. On this occasion the break down was forty percent from the medical school and forty percent from biosciences, with the school of chemistry, the school of pharmacy, the Welsh Development Agency, and the school of social science providing the remaining twenty percent.

INTRODUCTION

> *"Life is a foreign language; all men mispronounce it"* [Morley as cited in Campbell 2004].

This chapter highlights the role multi-skilled researchers and informal spaces play in aiding the stabilisation of scientific research fields by discussing five *matchmaking* mechanisms. The current climate of contemporary science is often interdisciplinary and involves actors, technologies, knowledges and practices from heterogeneous disciplinary backgrounds. These *assemblages* need to be translated across traditional disciplinary boundaries in order to forge the new collaborations that are the heartbeat of the new biology. In this chapter, I stress the importance of (i) innovative architecture, (ii) face-to-face contact and (iii) boundary-people as media for translating and transferring [tacit/craft-[82]] emerging scientific knowledge between different academic disciplines. In turn this helps to produce new interdisciplinary collaborations and promotes inter-boundary comprehension. Within the research areas of bioinformatics and proteomics these techniques have certainly proven indispensable when attempting to make sense of new science(s) and bridging the certainty/uncertainty precipice.

The previous chapter discussed the position of bioinformatics and the roles of the bioinformatician and bioinformaticist. In this regard it analysed the creative input that bioinformaticians have on omic science and in particular on the high-throughput technologies that they produce and deploy. The creative and technical competence they embody aids scientific knowledge-transfer from researcher to machine or computer algorithm and back to the researcher again. This position includes not only constructing the new technologies, but also the ability to interpret any results generated from them. As Dr. Dennis explains this role also requires researchers to be knowledgeable in more than one traditional discipline:

> "I think there are really three different levels of informatics...One - you go to the Internet and paste in a different sequence...that is relatively straightforward. Then being able to understand the power of statistical analysis, and I think there you need a statistician in place. And then there

[82] I have inserted the words tacit and craft to indicate that some knowledges may contain tacit and/or craft knowledge.

> is the writing of software. I can do the first, I can do some of the second and I have absolutely no interest in writing software at all...It is something you just learn...I have seen the informatic thing evolve and just grown with it."
> **[Dr. Dennis: Lecturer in Genetics and Molecular Biology Research]**

Dr. Dennis believes that to do the second level analysis of bioinformatics that he describes above you must be able "to understand the power of statistical analysis". In saying so he introduces the importance of a new breed of specialists to big biology projects; *'the statistician'* - an actor who is able to interpret data generated from technologies and then represent them as biological knowledge. This suggestion also introduces the concept of a *statistician* working with a *biologist* on an interdisciplinary research project (Chapter Seven). In this chapter I continue to focus on the theme of knowledge, but on this occasion I begin by concentrating on interdisciplinary languages, before analysing big project collaborations through the idea of communities. More specifically the chapter (i) focuses on the role of *matchmakers* (researchers who often head research projects) and illustrates the fundamental part they play in collating all types of knowledges, (ii) describes some of the techniques that are being employed by matchmakers to improve interdisciplinary communication and linguistic understanding, and (iii) tentatively suggests the emergence of new type of collaborative biology called *permodern science.*

In summary, the main argument in Chapter Eight is that *propinquity* and *scientific matchmakers* are integral to emerging scientific knowledge-transfer, particularly in research areas that attempt to generate knowledge through new interdisciplinary collaborations. This all suggests that there is a reconfiguration of the biological vista based on trust, proximity, reputation and size.

CUTTING-EDGE COMMUNICATION

One of the greatest demands of interdisciplinary research is getting assorted experts from heterogeneous backgrounds to communicate and comprehend each other. This challenge is best illustrated in omic biology where numerous knowledges, each with their distinct individual language, need to be decoded,

translated and *'matchmade'* by researchers attempting to create ordered knowledge out of complex riddles.

Today omic biology is a global institution comprising an international community (Beaver 2001). In line with Christopher Morley's quote (p205), due to new imaginations of what biological science can achieve, research in this area has almost become research on *life* itself. What I mean by this is that omic biology's boundaries have expanded to such an extent that in the twenty-first century it is home to many more diverse languages than its traditional smaller and more reductionist biological counterpart. The languages the new biological landscape contains include not only the multi-national, national and local cultural languages such as English, Cantonese or Catalan, but also heterogeneous scientific bordered languages such as the languages of chemistry, of biology and of mathematics highlighted in Chapter Five. As Beaver (2001) comments: "no one region, nation, or civilization remains the center of creativity and activity for long" (p365), and biology's extension means that even more new geographical hubs of activity are beginning to emerge. As a consequence interdisciplinary communication and comprehension have become essential skills if actors are to generate new biological knowledge exploring the functioning of whole systems. To successfully accomplish this venture however, it may mean actors reorganising, and sometimes exonerating, past ways of performing science. Below is a clear example of the problems that exist as a result of different traditional disciplinary languages. Dr. Francis is commenting on terminologies in the RESID[83] database:

> "Oh yes. Well the most historical thing is phosphohistidine. The chemists would call something one (1) phosphohistidine because the particular position the phosphorous was on the histidine range but the biochemists would call the same thing three (3) phosphohistidine and the problem is, of course, is that the three (3) phosphohistidine would [be] call[ed] something else so they would be calling the compound the other name and it was very confusing. He said (1) phosphohistidine, is he a chemist or a biochemist? And so finally in the late 1980s they came up with a new way of naming these. They called it prospotiline, but nobody knows what prospotiline means so you have to go and look it up, or you look in the RESID database and I put in a little drawing. This is prospo or one (1)

[83] Hosted by the EBI, the RESID database is a comprehensive collection of annotations and structures for protein modifications.

> phosphohistidine if you are a chemist or three (3) phosphohistidine if you are a biochemist. So all of this is explained in the RESID entry, and we don't have to go into any detail in SWISS-PROT, or in the literature table, we just say prospophosphohistidine – oh fine."
> **[Dr. Francis: Biotechnology Research Scientist]**

In this interview Dr. Francis illustrates the different labels of one and three 'phosphohistidine' that existed in chemistry and biochemistry. He argues that their coexistence was a problem of autonomous traditional disciplinary categorisation with one label meaning something in one discipline and another label meaning the same thing in another. According to Dr. Francis the problem was not incommensurate and could be overcome by re-naming and re-categorising the nomenclature to create a new one that can be understood by all disciplinary camps. This was achieved by replacing the terms one and three phosphohistidine and creating the new term 'prospotiline'. The important role that Dr. Francis performed was to *matchmake* the new classification to the chronological nosography of all its original autonomous disciplinary labels. This re-naming of old disciplinary nomenclature is one of the challenges for interdisciplinary work, but one which Dr. Francis believes is evidently doable and can be sorted through efficient language categorisation. Both community-based terms were *tinkered* with and a new *standard interdisciplinary term* was created.

This described account is an example of the identification/juncture phase illustrated in Chapter Six, where the construction of standardised terminology is required in order to regulate interdisciplinary research. The process may be time-consuming and bureaucratic, but it is not overly taxing since both the chemists and biochemists in this example are speaking similar type languages. The only real difference between the two communities was that they have different terms to explain the biological structure under scrutiny. Consequently, through the construction and implementation of a new interdisciplinary vocabulary, Dr. Francis takes on the role of a *'standard language matchmaker'* matching the newly created term with old autonomous disciplinary understanding in the RESID database. The construction of the vocabulary is one step towards the composition of a common ontology that may aid communication between all the communities involved in this interdisciplinary research area.

Despite Dr. Francis' achievement, what this chapter demonstrates is that a higher level of matchmaking is often required in emerging research fields since inter-group communication is about sifting through the different meanings, significance and interpretations of various sub-communities in order to create a unified science. Galison (1997) is important here. He believes that science is made up of a kaleidoscope of diverse and constantly changing languages and practices. But despite variances in the interpretations of what objects and languages may represent, heterogeneous groups can often come to some sort of conformity:

> "Two groups can agree on rules of exchange even if they ascribe utterly different significance to the objects being exchanged; they may even disagree on the meaning of the exchange process itself. Nonetheless, the trading partners can hammer out a local coordination, despite vast global differences" (Galison 1997, p783).

Galison's focus is on two or more distinct groups and how they go about communicating. His belief is that variances are sorted out on the borders of science, in what he calls *trading zones*. It is at these boundaries that different ideas and objects are exchanged and new inter-languages such as biochemistry are created. This is an interesting disclosure, however, in the case of the 'new biology' alternative interpretations of meanings are not just restricted to the borders and can also be evident within the same discipline or the same community. When this occurs, it leads to a type of *intralingual* translation problem that Jacobson describes. Jacobson (1959) states that even within the same discipline the translation of some terms may suffer from problems of equivalence, and if this is the case, even synonyms cannot capture the adequation of the word. An example of this type of *intralingual* problem is described by Dr. Andrews, who tells how a term may be used in different contexts within the same setting:

> "Even in micro-biology or molecular-biology the same term is used in different contexts. I mentioned chimeras earlier. I have seen chimeras used in completely different contexts within biology and that is before you move into computer science or more broader areas. I think it is just inevitable as we are constantly creating new terms and new worlds...You

cannot prescribe the English language and you cannot prescribe scientific language."
[Dr. Andrews: Lecturer in Biosciences]

The extract from Dr. Andrews suggests that shared meaning is difficult to attain even when actors reside in the same disciplinary camp and have been trained in similar ways. The challenge interdisciplinary *matchmakers* face though, is to help create shared meanings between heterogeneous disciplinary camps who each may have diverse individual histories and understandings of what certain words mean. This is an important condition because shared meaning and shared understanding is a prerequisite when communicating and comprehending knowledge. Jacobson (1959) called this type of transfer, '*interlingual translation*'; a type of between languages translation that Galison is describing.

Within the worlds of proteomics and bioinformatics interdisciplinary research becomes further complicated however, since disciplinary[84] languages are not the only languages contained within omic biology that need to be sorted by *matchmakers*. To add further complexity omic biology also contains the intricate language of the genetic code (As, Cs, Ts, and Gs) and the binary language of the computer machine based on 0s (zeros) and 1s (ones)[85]. In some senses these are the more complex and powerful languages and may mean the role of *matchmaking* within omic biology becoming increasingly more difficult than in other interdisciplinary arenas (Chapter Seven). The strength of this type of binary language is illustrated below in the dialogue from the film, 'The Core' (Dir. Amiel 2003); a science fiction film in which six scientists have to drill down to the centre of the earth to set it spinning again. Taz 'Rat' Finch is a computer hacker that the government has employed in this project and Dr Zimsky is a world famous physicist.

Taz 'Rat' Finch: How many languages do you speak?

Dr. Konrad Zimsky: Five, actually.

[84] The spoken language is entangled and embedded in the language of the specialism.
[85] This movement has been coined e-science (RCUK 2008).

> **Taz 'Rat' Finch:** I speak one. One zero one zero zero. With that I could steal your money, your secrets, your sexual fantasies, your whole life. In any country, any time, any place I want. We multitask like you breathe. I couldn't think as slow as you if I tried.

As the dialogue illustrates, although the binary language of the computer may seem simple and somewhat inconsequential on the surface, when utilised as a sequence it opens up numerous opportunities. 'Rat' may only speak one language, and a silent language at that, but in certain circumstances and for particular requirements, understanding that language is much more powerful than understanding five spoken languages. The same could be said to be true in omic biology where *demiurges* utilise the silent (non-hearing) language of the computer machine in order to interpret and analyse the masses amount of data generated (Chapter Seven). Dr. Campbell comments on the importance of bioinformatics:

> "Proteomics could not exist as it does today without bioinformatics. It would be a much lower throughput discipline."
> **[Dr. Campbell: Lecturer in Bioinformatics and Proteomics]**

Considering there is such an array of dialects and their contained skills, how does a mathematician or informatician communicate with a biologist on an omic science project? Or again using the idea of languages, how does a mathematician or informatician whose traditional training and expertise has been to learn the language of mathematics and binary numbers understand As, Cs, Ts and Gs, and how does the biologist who has been brought up on a vocabulary of genetic codes learn to understand the language of mathematical algorithms? Additionally, are all these languages and interpretations as commensurate as the story of the 'prospotiline' solution *matchmade* by Dr. Francis? The rest of the chapter will address these questions by concentrating on interdisciplinary collaborations in a period that I tentatively call *permodern science*.

INTERDISCIPLINARY RESEARCH AND THE APPEARANCE OF *PERMODERN* SCIENCE

As discussed in Chapters Two and Six, systems biology can be defined as the holistic top down view of biology. It is a *method* of interdisciplinary research that intends to look at the system as a whole by studying how each particular pathway

interacts with one another. This approach is the epitome of big biological omic science because it is a challenge based on a network-typed analysis of whole-system biological relationships within the body. In Chapters Two and Three I suggest that this could also be seen as a paradigm shift within biology with a swing from piecemeal science to a type of composition science. For systems biology to function however, I maintain that there needs to be an equally efficient social network integrating the relevant disparate actors that are required to analyse whole biological systems. As this chapter illustrates, this network is pulled together and bridged by *matchmakers* whose challenges include creating common ontologies, encrypting codes, fitting any of the missing pieces of the jigsaw together and organising all the heterogeneous knowledges generated. These challenges are a response to a possible stage within science based on interdisciplinary collaborations within the academic setting that I call *permodern science*; a fusion of the words permeate and modern.

Sociology has been consistently fascinated with eras, epochs and ages and specifically the transition from (i) ancient societies to feudal societies, from (ii) feudal societies to modern societies and from (iii) modern societies to post-modern societies. However this interest has often been in relation to work and culture and the movement into and from industrialisation and post-industrialisation and fordism and post-fordism. But this fascination is also apparent within science studies. For example Popper (1959), Kuhn (1962) and Shapin's (1996) work has evaluated revolutionary changes and stages in science; a term that denotes epochs. While Latour (1992), Beck (1992) and Functowicz and Ravetz (1993) have explicitly discussed the relative stages of modernity. Functowicz and Ravetz (1993) build on Kuhn's seminal work of analysing modern science by introducing the concept of post-*normal science* to the literature, which develops Kuhn's *normal science* ideas. For them post-normal science is the appropriate term to describe a science where "uncertainty is not banished but is managed, and values are not presupposed but are made explicit" (p740).

Whereas Functowitcz and Ravetz have concentrated on the practices of post-normal science, Beck (1992) and Latour (1992) have specifically been concerned

with the term 'modern' in modern science. In Beck's case this has been to criticise social science's commentary on social change by claiming that modernisation produces outcomes that the theory of modernisation has failed to recognise. In response to this Beck advocates the need for a more self-reflective modernity that is able to capture the risks, consequences and dangers being created in modern science; in essence asking for a more mature and accountable science. However, if Beck wants a more responsible modernity, Latour believes "*We have never been modern*" since he argues that our idea of modernity is founded on an artificial separation of nature and culture. For Latour modernisation is based on a *purification* process by which science is able to rise above nature by applying reason to certain situations. However Latour questions whether we have ever managed this, citing the example of Climate Change as a phenomenon in which nature and culture commingle as a hybrid form difficult to disentangle.

When talking to biological scientists however, there appears to be a new optimism that genomic science has created a new regime of truth. As Professor Llewellyn, head of a chemistry department, remarked in an e-mail sent to me: "We shall look at old-style science and laugh. We are on the precipice of something new". This e-mail was sent in the summer of 2006 and suggested that biological scientists believe we are entering a new scientific era; a period characterised by an underlying explanatory core (Hood 2003). This new type science is, in part, captured by Gibbons *et al.* (1994) who coin the terms Mode 1 and Mode 2 science. They argue that:

> "In Mode 1 problems are set and solved in a context governed by the, largely academic interests of a specific community. By contrast Mode 2 knowledge is carried out in a context of application. Mode 1 is disciplinary while Mode 2 transdisciplinary. Mode 1 is characterised by homogeneity, Mode 2 by heterogeneity. Organisationally, Mode 1 is hierarchical and tends to preserve its form, while Mode 2 is more heterarchical and transient" (Gibbons *et al.* 1994).

Gibbons *et al.* (1994) believe that the new production of knowledge has changed so dramatically in recent years that we need to distinguish between two types of science; Mode 1 science and Mode 2 science. They characterise Mode 1 as the

classical definition of science that is created in traditional research universities and ordered into a hierarchy of disciplines with physics placed at the top of the pyramid. In Mode 1 science there is also a clear separation between applied and basic research, with the whole institution of science having a certain mystique or aura that keeps it separate from society and what Gibbons (2007) describes as *on transit*. This is different to Mode 2 science where they argue that science is thoroughly integrated with society and with economics in particular. Furthermore in Mode 2 science the sharp distinction between applied and basic science is broken down and there are other non-academic ways of organising the production of knowledge such as Research and Development consultants, Non Governmental Organisations and Think Tanks. This new science is also contextual, adept at solving problems and utilises transdisciplinary collaborations as its mode of producing knowledge. They are careful to use the term transdisciplinary rather than multi or inter, believing that its semantic appeal implies a lack of respect for the old traditional institutional boundaries that Mode 1 science creates.

Building on the ideas of Mode 2 science by Gibbons *et al.* (1994), I tentatively and cautiously speculate about the development of a possible new rhetoric of change that may lead to a new era within academic science called *permodern science*. This concept of permodern science differs in subtle but important ways to the concept of Mode 2 science. Whereas Gibbons *et al.* (1994) concentrate on the ways in which knowledge is produced both inside and outside the traditional institutions of knowledge generation, the focus of this study has been solely located inside the traditional academic institution focussing on how biological scientists respond to both the uncertainties of everyday practice and to non-traditional challenges to authority. Within *permodern science* relationships are also more interdisciplinary rather than transdisciplinary and despite boundary permeations, there still appears to be recognised and institutionalised *scientific hierarchies*. This hierarchical relationship is emphasised in Chapter Seven where I discuss how bioinformaticians are struggling to get the recognition they believe they deserve within the biological community. It is further supported by Dr. Edwards who comments on the relationship between biology and computing:

> "It's the biology that is still the most important; the computer science just supports the biology."
> **[Dr. Edwards: Doctor and Lecturer in Molecular Cell Biology]**

In this sense I argue that *permodern science* within biology is characterised by:

> A period where interdisciplinary collaborations are central to knowledge production and traditional disciplinary academic boundaries are permeated by heterogeneous actors. Big science is beginning to dominate the field and knowledge is disembodied from actors and replaced in technologies and computer algorithms. E-mail and e-science is also part of the big science revolution in communication, in which texts become more proliferated, fluid and interpretable and scientific knowledge more instantaneous, asynchronous and non-attributable. There is a greater trust in numbers within biology as verification practices and new expertises need to be learnt. Consequently traditional expertise is changing and challenged as different, historically less recognised expertise is acknowledged, and yet at the same time these disciplines are still struggling to overcome some of academia's traditional hierarchical hurdles. The results of these new relationships mean that scientists are recognised within the scientific community more by their scientific identity rather than any specific expertise (the particular skill set that they bring), since in new emerging sciences scientists are often willing to bring their skill set (expertise) and contribute to the area, but are reluctant to identify themselves with it (see Chapter Seven).

Within permodern science, I maintain that there is also a greater visibility of the fractional scientist[86] (Price and Beaver 1966) and the appearance of the matchmaker. The role of the matchmaker is integral in order to understand and matchmake all the contradistinctive and diverse knowledges and expertises that are emerging. Despite the appearance of e-science, physicality of knowledge-transfer and face-to-face contact is fundamental in order to translate knowledge between communities and to assemble or create new knowledge out of complex labyrinths of ubiquitous data. Thus the permodern transformation of science is a brand new one where there is a new imagination of science. One in which Bertrand Russell's quote that "science may set limits to knowledge, but should not set limits to imagination" (2004, p26) is followed to the letter, and one where knowledge is managed, modulated and made sense of by multi-skilled matchmakers.

[86] A fractional scientist is a partially anonymous scientist who appears as a name on a journal article.

LOCATION, DISLOCATION, LOCATION

The integration of the WWW has been one of the key indicators of *permodern* science and one of the huge developments in science (Chapter Two). The new Information Age that the Internet has afforded science has been wonderfully captured by Manuel Castells (2000), who engagingly maps the trajectory of the journey that we have taken into (post)modernity. Within science the Internet has not only made new tools such as bioinformatics available and provided a platform to disseminate scientific information (Chapter Two), but it has also extended the potential locale of communication and collaboration among scientists. Beaver (2001) comments on the possible scientific futures of the Internet and e-journals:

> "There is a space and time for only a few limited and necessarily speculative ideas about possible future changes that may affect the form, quality, and nature of collaborative research in the future. In particular the expansion of the World Wide Web, and the growing number of electronic journals are likely to bring changes in research practice, which will be in turn reflected in the conventions of formal 'publication', whether singly or multiply authored" (Beaver 2001, p375).

It would appear that the Internet is one of the emblems of big science since it makes the world a smaller place and makes science more global by making geographical boundaries invisible. Yet in the specifics of my empirical case there is suggestion of an increased reflexive gaze on the problems as well as the solutions that the Internet may provide science. For example there is recognition that disembodied ubiquitous information must be vetted and regulated through (community-based) standardisation, and that the potential scope of a global scientific community may lead to dislocated and disenchanted researchers. Below is an extract from Dr. Cherry in which he recounts collaborating with a Brazilian scientist.

> "....I have just come back from a sabbatical in Brazil which was very nice but working with my colleague there for two months we submitted three papers and mostly wrote a fourth. Because you are there face-to-face it is much easier to thrash out anything, especially with language difficulties. You can just sort it out. Whereas by e-mail it just never gets done and other things come across your desk. So there is a great advantage to proximity but on the other hand the net and e-mail have opened up a lot in

> that I can work with her in some sense...However, we must remember that science is a communal activity."
> **[Dr. Cherry: Bioscientist and Molecular Biologist]**

Dr. Cherry champions the success of the Internet as a device that has aided communication and allowed him to collaborate with a colleague in Brazil. The Internet has extended the boundaries of scientific projects and has encouraged inter-continental collaboration. New possibilities have been created because of its ability to transcend traditional geographical and disciplinary boundaries[87]. However, there is a caveat to his assessment because he states that this extension should not be at the cost of face-to-face local contact as with "face-to-face it is easier to thrash out...language difficulties".

With the onset of the Internet there is the possibility that Dr. Cherry may not have to leave his office to conduct his research since he is able to communicate with researchers across the globe by typing on his keyboard or picking up his telephone. In this regard it is ironic that big science projects such as the Human Genome Project have paved the way for the possibility of very small science[88] that may involve one researcher, their room and their computer. Dr. Cherry feels that this type of engagement within science could be disastrous and might dislocate the researcher from his community since he suggests *science* is a plural term based on community activity. For instance, science is built on the rationale of being correct, and confirmation is often only achieved through organisational verification (Fleck 1979, Fine 2006); a process which needs community communication. Consequently we are beginning to see some of the changing practices that Beaver (2001) predicted.

Dr. Cherry continues the conversation by explaining how he recently submitted a grant application to work with a researcher from Cuba:

> "I have just put in a grant to get a Cuban scientist over who is a computer modeller of proteins but if he had already been working down the corridor I would have been more than happy to go bend his ear a long time ago...It

[87] The Internet could be viewed as a boundary object.
[88] Small in the sense of space and numbers.

> is often an issue that you don't know who you need until you know what they can do...seeing is believing."
> **[Dr. Cherry: Bioscientists and Molecular Biologist]**

Despite attempting to collaborate with someone on the other side of the world, Dr. Cherry confesses that if there were researchers available he would have collaborated with someone just as good in his own department. It is apparent that Dr. Cherry has seen and utilised the benefits that e-science has provided science, but still preaches the virtue of face-to-face physical contact, which among other things, he states, helps sort out any language and interpretation difficulties. Here is a suggestion of a movement towards *permodern* science. The Internet has opened up far-reaching opportunities that can extend collaborations and promote information dispersion. However there is recognition among respondents that it also needs to be regulated and managed, and above all should not be exploited by sacrificing face-to-face contact. For example e-services and e-science can extend networks and lead to increased mobility but they can also reduce the need for physical mobility through e-mail (Adam, Harris and Lewis 2002). Reduced physical mobility can lead to reduced face-to-face interaction, which in turn can create problems of communication. The problem, according to Dr. Cherry, is that issues can be negotiated and resolved face-to-face, whereas language difficulties can never truly be negotiated using a monitor or keyboard because different interpretative meanings cannot be observed nor debated. This all suggests that computing and communication do not necessarily go hand-in-hand. Schlossberg's statement is relevant here too:

> "True interactivity is not about clicking on icons or downloading files, it's about encouraging communication." [Schlossberg as cited in Weiners 2002].

It would appear that despite the Internet making mass scale communication a possibility where boundaries become invisible and imagined, physical and visibly authenticated spaces are fundamental to scientific research. They are still the legitimate and agreed spaces that can stimulate creative communication between scientists, and as Dr. Nielson argues, the most traditional of these spaces is still the formal academic conference:

> "Conferences are a major network point for several reasons. One is that sitting here I have got exam marking to do and exam board things to do and emails coming in. It is very difficult to sit down and let your mind not necessarily wander, but give it a bit of free time to think of things to do. At a conference you have that. You have a mix of people who come up and say 'that is an interesting thing you said, I wondered if...' or you see something they present and you think 'ahh perhaps we can put something together'. That kind of interface at a conference is extremely important, particularly in big science."
> **[Dr. Nielson: Reader in Molecular Haematology]**

Issues of physical space are important for Dr. Nielson since they can provide a creative environment that encourages stimulating engagement. This environment may not just be restricted to the formal and funded spaces of the conference, however, as informal spaces may be just, if not, more important (Collins 2006). In the extract below, Dr. Cherry bemoans the lack of a departmental tea room, which he believes may have enhanced his opportunities to find collaborators within his own department:

> "One of our long term complaints in this department is that we lost our tea room and it sounds very trivial but by in large it means we don't bump into other people down the corridor. If there is somebody I know to speak to about such and such I go and see them, but you don't get that casual contact [when] you find out somebody new has come and used to work on Malaria and pick up new ideas and things like that. The casual contact is also very important and is easily lost....I have had the situation before now where I have ended up collaborating with somebody overseas and later finding out that there was somebody in the same building that could have done the same thing for me, but I was unaware of their presence and came across the other person first and set it up."
> **[Dr. Cherry: Bioscientist and Molecular Biologist]**

Dr. Cherry explains that the loss of something seemingly as trivial as the department's tea room had a detrimental impact on intra-departmental collaboration. This type of café based culture is also discussed by Thrift (2006), albeit on a somewhat larger scale, who discusses the performativity of new bioscience buildings such as the systems biology co-location campuses. He stresses their porous nature and transparency in creating an '*innovative incubator*' that aids knowledge creation. Beyond this, Stephens, Atkinson and Glasner (2007; 2008b) show that the Centre for Life building in Newcastle has a further form of performativity. They state that the *performative architecture* of the

building is one of a socially transparent science where family and school trips can interact with cutting-edge science based establishments, aware of the contemporary models of public understanding of science. These physical spaces encourage *informal interaction* and *casual communication* between individuals. To quote Thrift (2006) again: "the aim is to make architecture more effective by making it more performative" (p292). In this sense, tea-rooms have the advantage of being non-hierarchical and non-exclusive spaces that encourage performance. These examples suggest that physical spaces can also become a type of *matchmaker*, providing an engaging environmental entity that pairs different researchers together and encourages the exchange of ideas and the nurturing of concepts. Following on from Thrift (2006), I have categorised this matchmaking mechanism - *innovative incubator*. Support for this was also found when I presented some of my findings to a group of scientists at Hinxton. One of the conference attendees described a similar incident in which they lost their smoking room and with it lost a similar informal local space where social interactions regularly occurred.

SCIENTIFIC SPEED-DATING

I have already described the roles of (i) the *standard language matchmaker* and (ii) the *innovative incubator*. The third type of matchmaker that I discuss in this chapter is the *manipulative matchmaker*. Whereas standardised and controlled vocabulary aids communication and mutual-understanding, and local physical spaces can provide responsive environments, unless they were built with a clear purpose of collaboration in mind such as the systems biology buildings they do not intentionally go out and promote collaboration between disciplines. Instead, other more forced techniques need to be employed by matchmakers to encourage dialogue when actors are not so forthcoming. In this case the *manipulative matchmaker* physically attempts to promote interdisciplinary research by positioning actors in the same locale. The first example of this type of matchmaking is what has commonly and colloquially become known on the dating scene as speed-dating. The below quote is from Dr. Jamieson who explains how he was asked to run a life science day in order to encourage

interaction between chemists and biologists. Due to its length the quote is split in two:

> "I have been involved in the last couple of years in trying to get the chemistry and biology life science interface bit up and running... Every single one of the collaborations that I have ever had with somebody I have had a personal relation with that person first and then the science comes out of it afterwards, because there is a degree of trust that you need with the other person and it is quite an intimate thing. There [are] a lot of negatives with collaborating with a person, if you share data before it is published they can easily take your data, run away with it and go and do your experiments...The idea that you can get people who work in the same field, sit them in a room and they will go and write a grant together, no they will not do that."
> **[Dr. Jamieson: Professor in Molecular Biology]**

In the first part of the extract Dr. Jamieson stresses the importance of developing social relationships and trust with potential collaborators. Collins (2001) supports this point in his case study of the quality factor Q of sapphire. In the article he advocates the need for face-to-face personal contact in order to build trust between scientists and potential collaborators.

In the interview Dr. Jamieson continued his response by describing how he has actively promoted interdisciplinary social networking:

> "...I won't go and write a grant with somebody who I have just met two minutes ago even if somebody told me we overlap...Therefore getting a social network where people from both sides learn each others language so they can learn to communicate in the first place and, as part of that the directorate people will form a relationship so you will know whether you can trust that person and you would like to work with them. You are not going to want to work with a psychopath...and there are quite a few, especially in chemistry. There are a lot of Mozart personalities[89] around I can tell you...I went to a cocktail party and there was a hundred single men all standing around with two metres space around them with beers and no-one speaking to each other, it was ridiculous...This was an idea that we came up with, bizarrely at Hinxton, where we hosted one of these chemistry, life sciences interfaces at the lecture theatre there. If you want to get together fifty biologists and fifty chemists and forge those links, knowing that chemists have some of those life skill issues how do you do it? I rather jokingly suggested speed-dating.

[89] My assumption is that the respondent meant Beethoven personalities.

> That was a facetious comment that was taken on by the BBSRC who overheard. So we run a speed-dating at a Royal Society of Chemistry life scientists interface meeting...I was mortified by the idea but it worked really well. What you do is put all the biologists with green named badges in a semi circle in a horseshoe. Then in the middle in a slightly smaller horseshoe facing them, we put the chemists with a different colour badge on and a whistle goes every three minutes and the inner circle people click round one person and rotate. So you have three minutes to say who you are, what you are interested in and the other person tells you what they are interested in and you find out if you have got anything in common and you make a note on a sheet who you like. It is just like speed dating....It was brilliant as it enabled me to run through a room of fifty people and go 'yeah you, you, and you I have got something with, and you lot are perfectly friendly but actually we don't overlap' so I am going to have conversations with you. I found two people who were working in and battling with what I was interested in that I wouldn't have spoken to otherwise and had no idea that they worked on that and now actually with one of them I have got a £450,000 BBSRC grant...That sort of forced social network would be a very good idea."
> **[Dr. Jamieson: Professor in Molecular Biology]**

In the second part of the quote Dr. Jamieson explains that after a failed attempt to get heterogeneous actors to interact with one another he came up with the whimsical idea of running a speed-dating game to encourage communication. To use his words, the initial suggestion was a "facetious comment", but in practice, this worked better than he would ever have imagined. The concept succeeded in encouraging two camps to communicate and collaborate with each other. In fact one outcome of the networking technique was a £450,000 successful collaborative grant. In this case the traditional social networks and social cliques that developed on the back of traditional disciplinary backgrounds were fractured using a novel social networking technique and a new interdisciplinary network/collaboration was created. Granovetter's (1973, 1983) work in this area states that this innovation may lie in the weak ties of existing cliques; ties between what he might determine as acquaintances rather than friends:

> "The macroscopic side of this communications argument is that social systems lacking in weak ties will be fragmented and incoherent. New ideas will spread slowly, scientific endeavours will be handicapped, and subgroups separated by race, ethnicity, geography or other characteristics will have difficulty reaching a *modus vivendi*" (Granovetter 1983, p202).

Thus, it appears the scientific speed-dating technique is one way of discovering these weak ties and finding actors willing, albeit with a little persuasion, to break their existing disciplinary ties in order to create new interdisciplinary ones. Support for this was found when another interview respondent described a similar successful story about a speed-dating event. Dr. Morris was asked how he is able to collaborate with biologists. His response follows:

> "It is pretty difficult. Quite a lot of the time it starts because they have got some mathematical or statistical question so they e-mail one of us and we pass it around. That is historically how it has been done...Although the University is trying out different modes now and so we had speed-dating. There was a day they got a lot of people together and the people who went said it was very good and they really enjoyed it."
> **[Dr. Morris: Professor in Statistics]**

It would seem that the speed-dating experimental device that uses techniques from other social worlds, and which actively encourages interdisciplinary communication is one of a number of pro-active communicative networking techniques that matchmakers may employ. It is also a mechanism that is being funded by research councils. The chapter continues by describing another technique which uses more subtle techniques than this type of forced social networking.

PIDGINS, PICTURES, CREOLES AND CARTOONS

A fourth matchmaking technique I describe is employed by more *diplomatic matchmakers*. In this case the researcher may be asked to deliver a presentation to interdisciplinary groups from heterogeneous backgrounds. The challenge is to get each distinctive group to understand the presentation at a sufficiently high level so that it was useful to them. Below is an extract from Dr. Hardwick who was asked to give seminars at the European Bioinformatics Institute (EBI). The EBI is an organisation that claims to be at the forefront of interdisciplinary omic research (Chapter Two).

> "...I think it gets down to the level of, for example, seminars and that is one way that scientists really get together and they really communicate. And so over the years I have been at some really bad ones from both camps. The biologists can stand there and hear a seminar, which is very

> difficult to understand unless you really are working in that field, and the same can be true with the computer people, the bioinformaticians. When I was working at EBI we used to go to seminars every week...and quite often even people in my own group would stand up and put slides up that might as well have been a page out of a scientific journal. They were so detailed that you couldn't read them, you couldn't see them and you couldn't understand them and so people on the same project may not understand the detail of the level of the work that was going on the screen. To me that is a complete waste of time and everyone shuts off and nobody appreciates it. The only impression it gives you is the level of work that they are doing is very complicated but what has anybody learned from that? Not a lot. So when I was there and just before I left, I decided I would try and do a talk that would try and bring together the biology and the informatics together in one talk using the work that I have done as an example and relating it altogether with the molecules on the screen. It was really simple blobs on the screen and how that was put directly in a box in the database because I think this is a really big problem. At the end of it I was really pleased because I had people from both camps come along and say that they understood it and I thought it was quite an achievement for me to do that. But it was a really really simple talk. So I think you need to get people together and get them to try and understand each others world by giving really simple talks actually."
> **[Dr. Hardwick: Molecular Biologist]**

In the interview Dr. Hardwick advocates the need to provide accessible information when delivering presentations. This necessity is heightened when the presentation is to be delivered to actors from heterogeneous backgrounds that might understand varying specialised languages. In her answer she asks the rhetorical question, does anyone actually learn from information they do not understand? Her answer is quite predictably - no! Nonetheless she continues to explain how she solved her particular problem in an experimental seminar she delivered. In this story the solution was to use diplomatic techniques and to re-structure the complicated, specialised and jargonised languages of the sciences in order to create a more simplified generic language that could be understood by all: "...[by] get[ting] people together and get[ting] them to try and understand each others world by giving really simple talks actually". I pressed Dr. Hardwick on this point and asked what the common feature was that enabled the two camps to understand each other. Her response follows.

> **Interviewer:** But I am interested in terms of what you think is the common ground when you are doing this talk?

> "It is probably both [diagrams and simplification of language]. It is probably about making it simple enough for people to understand because I think there is a terrible tendency, it maybe in every field but definitely in science, that you have to feel that you justify your existence and in doing that people's default way of doing that is to try and impress their audience with huge amounts of work. But huge amounts of work, ok we all know everyone works hard, but again we don't learn anything and if we can't analyse what this huge amount of work was, we have learnt even less."
> **[Dr. Hardwick: Molecular Biologist]**

Dr. Hardwick critiques the performative roles of scientists. Rather than exhausting their energy presenting a formal picture of themselves by regurgitating their accomplishments and displaying their linguistic skills: "you have to feel that you justify your existence", she believes that their focus should be channelled into successfully transferring specialised knowledge from one camp to another camp and visa versa. In her experience this can be achieved in two ways.

The first is the development of simplified interdisciplinary languages or what Galison (1997) might have called jargon, pidgin or creole[90] languages. These intermediate languages work by stripping away any disciplinary bias. An example of this is provided in her account where she de-jargonised the presentation to produce a language that was neither biological nor informatic in nature. Using this technique the language embedded in a discipline becomes unpacked and simplified so that neither language is valued over the other. Thus the informaticians do not have to feel as if they have to learn the biologists' language, and visa versa. This diminishes any power imbalances that might emerge if only one camp had to learn the language of the other. Instead commonality is found in simplicity and simplicity found in commonality.

The second technique that emerges from her account is to utilise drawings and diagrams. The presentation of this new pictorial language, rather than not emerging from either camp (as in the de-jargonised example), actually emerges from both camps. Drawings and diagrams have traditionally been used in both biological and computing settings. Thus it could be argued that both camps are familiar with the language. If this is the case it means that the audience at her

[90] The simplest of the inter-language is called a jargon, more complex is the pidgin, while creoles are the creation of completely new languages.

seminar are able to get a good grasp of the content of the presentation, and once more she does not have to prioritise one discipline over another. Instead diagrammatic communication has sufficient overlap with both traditional disciplinary languages so as to act as a bridging device. The result was that "people from both camps...say they understood it" (p224). Moreover the pictorial language used was not sophisticated: they were just 'simple blobs on the screen' (p224). According to Mulkay and Gilbert's (1984) scientists prefer the simplified, abstract diagram as a representation of a process or an organism, since this type of drawing is not intended to be too realistic, and therefore is often less false than those that do. They are just representational pictures that can aid understanding and break down barriers: "pictures are working conceptual hallucinations. Nothing limits you when you make a picture" (Mulkay and Gilbert 1984, p156).

Using this diplomatic technique the diagram and drawing device fulfilled the role of a boundary object by entering multiple worlds and assisting to reduce the epistemic gap that existed. In contrast the construction of a new simplified language creates an interim, evanescent world that fits in between the existing worlds and temporarily introduces new modalities that help inter-boundary comprehension[91]. This type of communication has interesting parallels with debates about how science should be communicated and delivered to different lay publics.

BOUNDARY SHIFTERS, MATCHMAKERS AND CODEBUSTERS

The term *matchmaker* is similar to the idea of *boundary shifters* coined by Pinch and Trocco (2002). Pinch and Trocco use this term to describe actors who shift from one world to another and when doing so produce a change in one of the worlds. Their description of the term is illustrated below:

> "Not only do people change identities, transgress boundaries, and move from one world to the other – say, from engineering to music – but they also apply the knowledge, skill and experience gained in one world to the

[91] The construction of standardised terminologies is a very similar technique, albeit usually more formal and stable.

> other. Thus, a Bob Moog morphed back and forth between his engineering world and the world of musicians and in the process he transformed the synthesizer. We call such people 'boundary shifters – people who cross boundaries and in so doing produce a transformation'. For an organisation successfully to innovate, it must allow for such boundary shifting. Salespeople would seem to be quintessential boundary shifters" (Pinch and Trocco 2002, p314).

Pinch and Trocco (2002) coin the term *boundary shifter* in the context of the development of synthesizers and samplers. The term *matchmaker* is slightly different though since matchmakers do not necessarily continue to shift between different worlds. For instance in the Pinch and Trocco (2002) example there are the worlds of engineers and the worlds of musicians and both the worlds are stable settings. The matchmakers portrayed in this chapter however, live in a new interdisciplinary setting (for example the proteomics setting) that has been, or is being, created between two old worlds (biology and computer science). Moreover, those old traditional worlds are now contained within the new setting as sub-worlds. The job of the matchmaker is to translate across the sub-worlds and use all the knowledges and languages that experts in those areas bring to the new setting in order to create new interdisciplinary knowledge. As such they do shift boundaries, but more so as trendsetters and transmitters who are expected to make sense of proto-boundary objects.

The final matchmaker I describe in this chapter employs the technique of shifting between sub-worlds to make order out of complex codes. I have called this type of matchmaker the *codebuster*. They are similar to the boundary shifters, but they do more than produce a transformation because as interdisciplinary scientific researchers, they are able to break language codes in order to improve mutual understanding and aid the stabilisation of the new setting.

The following is a lengthy but edited interview extract with a leading proteomics facility director in the UK:

> **Dr. Matthews:** "We have immersed ourselves in a few pretty large scale collaborations, mostly with colleagues in (University X). [It] is an interesting fact that with something like proteomics the best collaborations are the local ones, because of this...need to be able to communicate well

and sit face-to-face. With e-mail and telephone calls it's not always as successful".

Interviewer: "Can I just pick up on something? You're saying of... local collaborations... [that] one of the reasons why it's probably better is that you can speak to face-to-face. Can I ask why you think that that is important within proteomics as opposed to perhaps, you know, as you said, speaking on the telephone?"

Dr. Matthews: "Proteomics, the most important part of proteomics is the experimental design. If you get that wrong, you might have wonderful data sets at the end but they don't actually mean anything. To design an experiment effectively you've got to know an awful lot about the system...and this is one of the challenges for somebody in my position, you have got to be able to understand the biological question, understand what the samples are, understand what state the samples are, how they've been prepared...it's absolute paramount importance that you can translate into biology, so proteomics is very techniques based, it's...quite a physical science, it's all about big mass spectrometers and getting anonymous peptides to fly. It's easy to lose touch with biologists and unless you keep that communication link and you can understand what it is that they're trying to do and everything there is to know about the sample, then there's no amount of good backend technology that is going to make a good project..."

(Large gap as interview continues)

Interviewer: "...bringing it back to what you were talking about at the start, and perhaps one of the developments or one of kind of the changes ....between other kinds of protein research to proteomics then perhaps I would suggest it is perhaps more kind of, interdisciplinary work. [So] that you'll get, not just chemists, biologists but curators, and information scientists... has there been a...change in terms of communication with these different kind of scientists and how do you go about...communicating with...these different kind of people who... you said had different languages but then also different skills as well?"

Dr. Matthews: "With chemists it's not so bad and I think that's mostly, for me anyway because I did a fair amount of chemistry as a student. I wouldn't say I'm a good chemist but at least I've got a fighting chance with chemistry. The discipline or the two disciplines where I don't have a fighting chance at all, but I'm highly dependent on them are bioinformaticians and computer scientists and statisticians and mathematicians. And it's the latter set [mathematicians] that I personally have most problems understanding and making them understand me. So they've got no biology background. They've no idea why you want to choose a set of biological replicates as oppose to a set of technical replicates, in so much that you can get hold of one and not the other, you

know what I mean. And so you've got to try and educate them about the biology and they probably haven't done biology. I last did maths a long time ago...But I have to be able to design an experiment, knowing that I've got some sort of statistical power at the end of it otherwise there's no point in doing it. So they have to educate me to what it all means and what the statistical test means and what you're assuming as you're doing the tests and whether those assumptions are actually valid for your data set. They may well not be."

Interviewer: "So in terms of what you were talking about earlier, in terms of doing proteomics, you kind of need face-to-face interaction?"

Dr. Matthews: "Yeah."

Interviewer: ...you know if you speak to someone on the phone and someone's given you kind of a regression or some of statistical test but on the phone you're thinking, I don't have an idea, no, but is that an example [of that]...?

Dr. Matthews: "Yeah, no absolutely right. I don't think there's very much that you can do with a lot of these people without a piece of paper and a pen in front of you."

Interviewer: "Yeah. And - and could you foresee people having both the biological awareness and the statistical knowledge and the computer skills to be able to do this on their own."

Dr. Matthews: "Not one single person...It's so broad and it's also very deep...So I think within the field and this is true of any field, you have those people who have got depth but no width and those people who've got width and no depth and you need both and in many positions I think I need to be the latter, that I have a broad knowledge but [I] need experts for each bit."
[Dr. Matthews: Doctor in Biochemistry and head of a Proteomics Facility]

The dialogue ends with Dr. Matthews stating that proteomics is very broad and deep, and that in her position as a project leader she needs to have a broad knowledge of the new emergent world, but with specialised experts for each old autonomous world below her. What she is explaining here, and as has been described in the preceding chapters, is that proteomics has a strong pull to attract disparate actors, but that no one actor has both the biological and computing awareness. As these actors are able to keep their own established identities (Chapter Five) and bring into the field their own specialised expertise, then proteomics becomes an interdisciplinary field, (or as I have argued - a proto boundary object) containing a mixture of different languages and backgrounds.

In this situation Dr. Matthews states that someone involved in a proteomics project (usually the project leader) has to blend all the different languages together while gaining a certain understanding of all the differing sub-worldly expertise that it contains in order to help the stabilisation of the field and to propel the new project forward. What Dr. Matthews is illustrating, to use Collins and Evans' (2007) terminology, is that at least one person involved in a proteomics project has to be able to interact with all the different actors and all their heterogeneous languages at an expert level. It is her belief that this job should be performed by the project leader[92]. Consequently, Dr. Matthews must attain, in much the same way as Chapter Four has argued a sociological interviewer needs to acquire, the level of an '*interactional expert*'. As a reminder interactional expertise is the level of expertise that enables a person to interact constructively with other experts, even if they are not able to contribute practically to the field (Collins and Evans 2007). In this sense, Dr. Matthews must be able to interact with all the different sub-worldly languages and disciplines that have been attracted to a proteomics project at the *conversational* level.

The problem is one of communication then. How do you overcome the epistemic gap between mathematics and biology? Dr. Matthews once again advocates the benefits of physicality and propinquity, and believes that conversations can only be done face-to-face. She states at the beginning of the quote that the best collaborations are those that are done locally and implies that proteomics cannot really be conducted over the phone or over the computer (Internet and e-mail). As a biologist she needs to convey to the computer scientists and mathematicians working on the project what the biological problem is in order to make them more biologically aware. Alternatively, and equally importantly, she needs to gain an adequate grasp of the languages of mathematics and informatics. Thus this chapter has argued that boundary-people or *matchmakers* are required for proto-boundary objects to exist and to help them stabilise into boundary objects. The function they perform is to act as boundary-people; actors who are able to absorb all the different knowledges and then decode and merge them to produce a new form of knowledge (in this case proteomics). Biological, chemical, mathematical

[92] Collins and Evans (2007) often classify a manager of a technical project as having referred expert.

and informatic languages need to be combined and understood to create the language of proteomics. In this example Dr Matthews attempts to keep her own identity (that of a biologist) but attains interactional expertise in all the other sub-worldly specialities in order to become a contributory expert in proteomics. Despite admitting that she does not have a "fighting chance" with mathematics or computer science on her own, with other actors on board she believes she can be "educated" in the new field.

According to Collins (2004b), the idea of contributory expertise is "what you have if you immerse yourself in the culture in a full blown way" (p127). It is an expertise level *arguably* greater than interactional expertise in which you can actually practice and contribute to a specialised area. It is what a physicist is to physics and what an international cricket player is to cricket[93]. Although I argue expertise is socially constructed and socially substantiated, in the sense that it requires the recognition of other similar actors in a particular field, according to Collins and Evans (2007) scientists do have *real* expertise and skills. In this case Dr. Matthews, who defines herself as a biologist, but who works in the department of biochemistry, would be acknowledged within the social world of proteomics to be a contributory expert within it (she practices it). This expertise is real but still requires community recognition. Furthermore, as a matchmaker, Dr. Matthews has utilised interactional expertise in all the sub-languages of the interdisciplinary research project in order to contribute to the new emerging field[94]. In doing so she has broken the codes of the other languages. This is achieved through face-to-face interaction, immersion in the field and by developing a good rapport with the fellow workers[95].

Inter-personal and communication skills are seemingly paramount and help stabilise the social network that has been created in research projects. This is what Collins and Evans (2007) have also termed *interactive ability*. According to Jenkins (2007) this type of co-operation and collaboration facilitated by

[93] An interactional expert in physics maybe a social scientist studying the area, and an interactional expert in cricket may be a cricket commentator.
[94] In the case of Dr. Matthews, it may be argued that she has interactional expertise in mathematics and chemistry, but contributory expertise in proteomics.
[95] Dr. Jamieson described earlier than gaining a good rapport with potential collaborators is essential.

interactional expertise is also fundamental to the US tuna fishing industry since collective work will likely fail, if one or more of the parties do not have the interactional expertise to engage with the other (Jenkins 2007). She also states that formal training and educational qualifications are not the secret to attaining high levels of expertise when such complicated goals have to be achieved. Instead what she believes is required is an assortment of interactional and contributory expertise. This is a point I address further in Chapter Nine.

CONCLUSION

In this chapter I have advocated the need for face-to-face physical contact and performative informal and formal environments as ways to instigate and aid collaboration in new research fields. All of this points to the fact that matchmaking mechanisms help the stabilisation of new research fields, which may depend on a myriad of scientific languages. The collaborative techniques used to aid the translation of new knowledge were classified as:

(i) the standard language matchmaker
(ii) the innovative incubator
(iii) the manipulative matchmaker
(iv) the diplomatic matchmaker
(v) the codebuster.

The focus of these techniques were to (i) standardise languages and ontologies, (ii) create inviting spaces of work, (iii) provide collaborative social events, (iv) utilise diagrams and de-jargonised languages as ways of communicating, and (v) position boundary-people as leaders of interdisciplinary projects. The common measure that these techniques have with one another is that they aid the stabilisation of new research terrains. The devices help to *matchmake* knowledge, technologies, actors and practices from heterogeneous backgrounds in order to create the new interdisciplinary research areas in which they now reside. It is possible to represent all the heterogeneous spaces of interaction and collaboration in matrix form. In this regard, Table 8.1 illustrates the types of spaces where I found scientists interacting and collaborating. I maintain that this can be

separated into physical spaces that scientists have traditionally exploited and the more innovative virtual spaces provided in *permodern science* (Table 8.1).

SPACE	LOCAL		DISPERSED
	Formal	**Informal**	
PHYSICAL	**Co-Located Campus Seminars**	**Coffee/Tea Rooms Shared Buildings**	**Conferences Laboratory Networks Organisations PSI Standards**
VIRTUAL	**Computer Laboratories Computer Programmes Intranet and Email**		**WWW Chat and Blog sites Online Warehouses Reference Databases**[96]

Table 8.1: Spaces of Interaction Found in Post-Genomic Science.

Despite Table 8.1 recognising these different spaces of interaction, this chapter promotes face-to-face contact in newly emerging research fields[97]. This is not to say that in the future interdisciplinary knowledge may not be passed on via more virtual modes, but that physical social networking is fundamental in research areas in their infancy. The chapter has also shown how talented matchmakers aid the translation of new scientific knowledge by acting as both boundary breakers and boundary creators and by reconfiguring scientific relationships and biological landscapes. To this end the spaces of interaction where scientists collaborate are continually being created and re-created. The next chapter will continue the theme of interdisciplinarity by examining the role of new interdisciplinary degrees and education courses as forms of scientific stabilisation.

[96] I recognise that some of the spaces may be positioned in more than one column. For example the PSI standards may be placed in both the virtual and physical space.
[97] In these fields knowledge may not necessarily be tacit, but it is not explicit either.

CHAPTER NINE:

EDUCATING 'NEW CHAMELEON' SCIENTISTS

> *"A chameleon does not leave one tree until he is sure of another"*
> [Arabic Proverb].

INTRODUCTION

The previous chapters have consistently rested on a 'narrative' perspective. The data have been derived primarily from interviews, and I have been presenting scientists' accounts of the emergence of new 'omic' knowledge. In this chapter I parallel that by discussing the emergence of new scientists – that is, specialists in interdisciplinary fields. In common with the previous chapters, many of the data are derived from interviews. They are supplemented from materials derived from participant observation. While the latter may introduce a slightly different flavour to some aspects of the account, there is no fundamental incompatibility or incongruity here. The interview materials derive accounts that are, in themselves, performative. They are narratives of scientific development and disciplinary transformation. Participant observation generates data on similarly *performative* acts. Disciplinary identity is as much a performance than any narrative of scientific advance (Coffey and Atkinson 2002).

The subject-matter also parallels the substance of previous chapters. In discussing the emergence and stabilisation of new types of knowledge, and their stabilisation, we must not lose sight of the equivalent process whereby new social types or identities, are produced and performed. Peer-driven credentials are conspicuous indicators of how scientists identify themselves. Academic qualifications are some of the most sought after and respected of these types of community-based certificates. The creation of academic courses and the potential qualifications that can be achieved are also key markers in assessing the extent to which a new academic research area has stabilised. For example, if a degree course is available in a research discipline and working scientists attain

qualifications in that course then the research area becomes more widely identified by a larger cohort of actors, and hence more stable. This, in turn, leads to the establishment of clearly recognised academic roles (Ben-David and Collins 1966).

Chapter Nine investigates to what extent and to what level bioinformatics and proteomics are being filtered into formal academic teaching, and discusses how this new type 'omic' knowledge is taught and learnt. To this end, in the sense that it shows how omic actors are produced, the chapter marks a response to a call from Collins *et al.* (1998) for new interdisciplinary specialists within biology:

> "The HGP [Human Genome Project] has created the need for new kinds of scientific specialists who can be creative at the interface of biology and other disciplines, such as computer science, engineering, mathematics, physics, chemistry and the social sciences. As the popularity of genomic research increases, the demand for these specialists greatly exceeds the supply. In the past, the genome project has benefited immensely from the talents of nonbiological scientists, and their participation in the future is likely to be even more crucial. There is an urgent need to train more scientists in interdisciplinary areas that can contribute to genomics. Programs must be developed that will encourage training of both biological and nonbiological scientists for careers in genomics. Especially critical is the shortage of individuals trained in bioinformatics" (p688).

A considerable amount of science studies work has focused on the emergence, stabilisation and social construction of scientific disciplines and technologies (Bijker 1995; Hedgecoe 2001; Jordan and Lynch 1998). This emphasis is also apparent in the four previous chapters which analyse how proteomics and bioinformatics have emerged and begun to stabilise. In the previous chapter, the specific spotlight was on the problems that multiple languages have in affecting communication and collaboration between scientists in emerging interdisciplinary fields. The focus of this chapter however, is the process of knowledge-transfer as *learning* in an emerging interdisciplinary field. It tracks the translation of knowledge from *science teacher* to *science student* and analyses how the development of academic courses aid the stabilisation of new fields. Nowotny (2008) argues that 'transdisciplinary' teaching takes a great deal of time and effort from both actors and institutions, while Etzkowitz, Webster and Healy (1998)

claim that changes in academia are notoriously slow. In accordance with these claims, the questions I pose in this chapter are:

(i) how is someone educated in an emerging interdisciplinary field?

(ii) how is expertise constituted in these new research fields?

(iii) how is new knowledge filtered through the academic system?

The chapter therefore assesses how a new field is taught in the area of proteomics, how new skills are learnt in the area of bioinformatics, and how a new emerging field is endeavouring to stabilise and authenticate itself through academic degrees.

The chapter also emphasises the notion of the *'hybrid chameleon bioinformatician'* - a group of biologists who made calculated decisions five years ago to expand their career opportunities and learn new informatics skills. The use of the word *chameleon* is a reflection on how they have adapted their identity to change with the evolving environment they now inhabit. Many leading bioinformaticians still identify themselves within the boundary of biology (Chapter Seven), since they still believe this will provide them with the most epistemic authority (Gieryn 1999). Yet science is a flexible activity, and with the development of new skills and expertises this group of 'biologists' have begun positioning and preparing themselves for a new scientific future. Finally, the chapter highlights the important roles of *learning* and *training* in scientific research areas attempting to create a stabilised practice containing its own authentic experts. In this way the chapter begins by analysing how expertise is acquired in cutting-edge interdisciplinary fields originating from two or more disciplines[98].

ACQUIRING EXPERTISE IN EMERGING DISCIPLINES

In this section I tackle the question of skills acquisition and expertise in emerging disciplines. The notion of expertise has been of interest for various sociologists (Collins and Evans 2007; Giddens 1990, 1991; Jasanoff 2003), who have tackled

[98] This chapter has a slight change in tone from the four preceding chapters. The first half is written in a more reflexive style that utilises *fieldnotes* taken on the RSSDP course (see Chapter Four).

the subject from different positions. For example, Collins and Evans (2007) treat expertise as real and as an embodied skill, while Jasanoff (2003) believes that expertise is attributed and substantiated by socio-political factors arguing that:

> "expertise is not merely something that is in the heads and hands of skilled persons, constituted through their deep familiarity with the problem in question, [individual traits] but rather that it is something acquired and deployed, within particular historical, political and cultural contexts" (p393).

This is not to say, however, that Collins has not been interested in the relationships that scientists make in positioning themselves as a core expert. He argues that: "the picture is first developed during scientists' training and continues to develop as a result of their relations with colleagues and through their continued work" (Collins 1992, p142). Thus, his work has illustrated how *core sets* emerge out of core groups and how skills are acquired through group immersion and knowledge-transfer.

Skill acquisition has also been the interest of Ravetz (1971) and Sennett (2008) who focus on craft knowledge and craftsmanship. Sennett, in particular, illustrates how the delicate types of skill required to build cathedrals are now resurfacing in other area of work such as designing the computing software system, Linux. This type of knowledge and skill is a type of social capital that is laden with what Polanyi (1962, 1967) and latterly Collins (2001, 2007) would call tacit knowledge (Chapter Eight).

In *Rethinking Expertise*, Collins and Evans (2007) produce a model of legitimate practice and distinguish between ubiquitous tacit knowledge and specialised tacit knowledge. In the remaining part of the chapter, I illustrate some of the ways knowledge and skills are acquired in newly emerging research fields and ask whether sociological work in expertise now needs to focus more on how new experts are created[99].

[99] This is also an issue that Eriksson (2004) deals with in her doctorate examining the Pusztai affair.

THE INTERDISCIPLINARY COURSE: MUTUAL LINGUISTIC SOCIALISATION

In her book, *'Invisible Colleges: Diffusion of Knowledge in Scientific Communities'*, Diana Crane (1972) describes how scientists have several different types of relationship with each other. One relationship type is that between 'science teacher' and 'science student'. She states that within scientific research projects:

> "Frequently, collaborators are teachers and students. Even without formal collaboration, the teacher who trains a student often retains a close relationship with him in later years. In any case, the teacher's ideas and orientation toward the field are likely to leave their mark upon the student's perception of the field" (p41).

During fieldwork, I had the opportunity to observe and experience this classic relationship between teacher and student first-hand. I spent time on a (RSSDP) Protein Bioinformatics course run by a Research Graduate School (Chapter Four). In one of the sessions the science teacher, Dr. Campbell, put up a slide describing some of the methodological techniques used in the field of proteomics. At which point, he stopped and turned to one of the six post-graduate students attending the course and cajoled him: "Jonathan, you are a molecular biologist perhaps you can explain this technique better than me" (*Fieldnotes* 2006). On first reflection this did not seem peculiar and I believed it to be a simple case of a science teacher encouraging student participation. It soon became apparent, however, that this was different to traditional student participation encouragement, and that the dialogue was more to do with the construction and comprehension of knowledges in an emerging field. The science teacher was a protein chemist by trade and training: "I'm a biochemist and that's really where I come from[100]", and it was only with the emergence and development of proteomics that he had become re-introduced to molecular biology techniques he last learnt when studying. It was Dr. Campbell's belief therefore that Jonathan might be able to give a better account of explaining the technique to the class than he could. Indeed, after

[100] Dr. Campbell's comment that his background is really as a biochemist is consistent with the findings of Delamont, Atkinson and Parry (2000) who argue that even in interdisciplinary fields, people often self-identify in terms of their original discipline.

Jonathan gave a short explanation to the class, Dr. Campbell acknowledged that "it certainly enlightened me".

It was this moment, while sat in the room as an observing participant, listening to the dialogue between Dr. Campbell and Jonathan that I began to consider how newly emerging interdisciplinary research is exactly taught. In this scenario, the teacher had willingly allowed the power, if power is to be viewed as knowledge, to be transferred over to the student. This meant that the balance of power during a short period of the lesson had swung to Jonathan who answered the query eloquently, informing both the rest of the class and also adding to the teacher's knowledge base. This scenario begs some important questions: Who is the expert in the proteomics technique? Is it the *science teacher,* as portrayed in Crane's account, or is it maybe the *science student*? Or is it a combination of both teacher and student, and is this what is meant by a 'community' of knowledge. If the latter is the case how did Dr. Campbell know that Jonathan's testimony was correct?

Initially I found this technique of teaching quite novel since it was not the *standardised* and *routinised* way to transfer knowledge in an academic setting. The more I evaluated the situation though, the more impressed I was by the honesty and open way that the teacher wanted knowledge to be shared. It became apparent to me that teachers like Dr. Campbell who are working in new emerging fields need to learn as well. With this revelation, I realised that his actions could actually be viewed as a functional response to the knowledge labyrinth that now exists in big biology. Jonathan is a Ph.D. student, and has a strong molecular biology background, so why not ask the student help him to teach the class, I thought. The enormity of the challenge and the breadth of knowledge required to practice in omic biology means, as earlier chapters have illustrated, that a cohort of open, knowledgeable and communicative researchers are needed to make sense of information. As argued in Chapter Seven and as illustrated by Dr. Campbell here, researchers should not marginalise or degrade any potentially important skilled contributor who may help to make sense of the puzzle. Instead, they should encourage, embrace and utilise them in the best way possible. In this example, this is achieved through *linguistic socialisation* (Collins and Evans

2007). However this is not traditional one-way directed talk from science teacher to science student; instead, a type of *two-way mutual linguistic socialisation* translated and transferred in both directions from teacher to student and student to teacher. In traditional teaching environments the teacher passes on their knowledge to the student and, as Crane (1972) describes, the teacher trains the student. In this setting however, Dr. Campbell has helped create an environment, where he can teach students but also where his students can teach him. In doing so he has also invited the student to share the *centre stage* so that he can inform the class. This is the result of proteomics knowledge still being in an embryonic stage of development (*phase zero science*) where shared knowledge, shared resources and dual engagement are paramount to knowledge production. Interestingly, Dr. Campbell has to trust the comments of Jonathan, much in the same way that Jonathan would usually have to trust the teachings of Dr. Campbell.

THE ROLE OF THE SCIENCE STUDENT AND SCIENCE DEGREE

I have illustrated how expertise can be acquired in an emerging discipline through *two-way mutual linguistic socialisation*, and I have shown how the technique of *sharing the centre stage* has proved to be useful within an interdisciplinary teaching setting. In this section of the chapter I continue to focus on language socialisation, but this time in the more formalised setting of degree qualifications. While the previous section focussed on proteomics, the remainder of this chapter focuses on the development of bioinformatics.

When discussing the emergence of the new informational biology in the 1990s, and the increased role that informatics will play in interdisciplinary biology, Leroy Hood (1992) wrote:

> "How can more scientists from other disciplines be brought into these efforts? One approach is to create a new kind of biologist – mainly by establishing Ph.D. programmes in biotechnology that build bridges to other disciplines. Such programmes would select students who wish to major in one area of biology, such as molecular biology, and in another discipline, such as computer science. The student would have a mentor in

> each area and take appropriate qualifying examinations in each. The objective would be to choose, for example, a fundamental problem in molecular biology and then develop and apply a tool in computer science that could be applied to it, thus bringing computer science into biology through the students. This programme would develop inter-disciplinary scientists, those with expertise in biology and other disciplines and the ability to forge interdisciplinary collaborations" (p149).

As with the science teacher - science student story I presented above, Hood also focuses on the *'role of the student'*. Hood wants to find an answer to the question: "How can more scientists from other disciplines be brought into these efforts?" Thinking about a solution to this dilemma, Hood suggests that it is at the student level of the *knowledge chain* that actors need to be trained in developing the new skills and languages that are required for interdisciplinary translational research. Presumably, his interest at this particular level is because he believes that the future of the research is in the hands of the next generation researchers who will be central to the continual development of the new biology.

During fieldwork, I asked bioinformaticians a question about the background they had in either computing or biology as a precursor to the type of questions that Leroy Hood (1992) asks. The response from Dr. Griffiths below was a typical reply:

> "Well my background is really biophysics. So I did a joint honours first degree in biology and physics, so in one day, in the same day, I was doing quantum chemistry and physics and the next day cell biology. I didn't actually feel that the course mixed it very well. Then it got me into computing, I did a programming course and then my interest grew from there. Then I did a Ph.D. in biocomputing, we still didn't have the name 'bioinformatics' in that place...I had always been interested in computers. I had one as a teenager, so I could programme a bit even before I went to University. I just liked the logical side of it, writing programmes to solve problems, and my natural interest in the science was leaning towards the bioscience and biomedical science so the two came together I guess."
> **[Dr. Griffiths: Reader in Bioinformatics]**

Dr. Griffiths' extract was a standard response representing the process by which the majority of bioinformaticians in the study learnt the computing components of their profession. They explained how they had become interested in computers as children, often by playing around with their home PC and coupled with their

academic training in biology, they were able to combine the two skills when entering the new field of bioinformatics. It would appear then that many of the bioinformaticians in the study combine the skills they developed in their leisure time (outside of academic learning and training) with the skills they learned during their academic training. This micro example adds further weight to the Mode 2 idea of science put forward by Gibbons *et al.* (1994) in which non-academic skills and expertise are increasingly being recognised as science.

In the specific case of Dr. Griffiths I wanted to delve a little deeper, in order to find out about the overall skills and knowledges that are present at his research centre. I moved the direction of the discussion towards the role that education has played and continues to play in the areas of bioinformatics and biocomputing at his University. Dr. Griffiths gives his answer in the extended dialogue below:

> **Dr. Griffiths:** "If you look at the research levels, we are talking about Ph.D. students, Post Docs, Lecturers and Professors, whether we are a good model for the rest of the country I don't know, but we have a lot of research strength in both. So there are four or five senior Lecturers and Professors in the computer science department who would probably get away with calling themselves bioinformaticians, although some of their research is not biologically focussed most of it is. Likewise we have probably got about another ten people in the faculty of life sciences who would probably call themselves bioinformaticians in some way, but are not computer scientists. So we probably do have a very strong balance in both communities there, and I don't know how well we can compare ourselves to other universities, but I would say we are probably one of the better institutions at merging the two. So at the Master's level we have a lot of contribution from the computer science department to the Master's degree we run. We have a lot of joint research grants all the way up. We even have a nascent interaction at the undergraduate level where we have the joint degree programme."
>
> **Interviewer:** "[So] is it fair to say that with these Master's, Ph.D.s and even at the Undergraduate level that younger researchers are more likely to be trained in both the disciplines and so this knowledge [base] will increase?"
>
> **Dr. Griffiths:** "Well that is interesting because we have tried and this was in 2000 when we were a separate...I started up an undergraduate programme then and that essentially was one of its underlying philosophies really. I felt that we did need to educate undergraduates from day one, with both strands, with the computing side and with the biological side. With the idea that you generate people within three years

who are pretty competent programmers, maybe not as good as a full computer scientist, but had all the biological knowledge that you really needed in bioinformatics. But obviously there were going to be some areas where they wouldn't have the breadth that a biochemist or geneticist would do, but hopefully in pockets, they would have the same depth of understanding. And the same would apply for the computer scientists as well. Whether it has proven a success is arguable.

So we are now slightly moving away from that model, partly to do with administrative things rather than the pedagogical reasons for it. That was the ethos behind that, and if you compare that with the traditional Master's courses, and it is the same here and it is the same in [University D], and everywhere else that I have seen, you tend to get people coming in from one camp or the other... At [University D] my experience is that from the outset it was quite a good mix. Particularly because there was a strong influence from the computer science department they could persuade a lot of their decent graduate to this kind of study. Computer Science has seen a trend especially in the very recent years where five or six years ago they couldn't keep even their worst students, they were getting taken up straight into industrial posts, now they are finding even their better students can't get jobs. So they are looking at PostGrad courses and careers. But having said that, it seems that of late, the percentage of people, let's say from the bioscience background, which will probably include the chemical sciences, coming into bioinformatics or indeed chemoinformatics has increased. So there are more biologists coming in now [and] so you have different challenges then. You have got to teach them computing when they know nothing about it."
[Dr. Griffiths: Reader in Bioinformatics]

It is clear from the dialogue that Dr. Griffiths has attempted to follow a very similar model to Leroy Hood. In his research centre, Dr. Griffiths and his colleagues had focussed on developing a clear structure of learning in bioinformatics from undergraduate level through to Professorship level. It was hoped that this model would create an infrastructure that could generate a body of learning and knowledge which is missing in newly constructed research areas[101]. Essentially, the model would encourage knowledge filtration through all the different levels of expertise and experience, and despite stating "we are now slightly moving away from the model", he claims that due to the reduction in the number of industrial posts, today there is a larger cohort of students trained in computer science remaining in academia. Subsequently, he believes that there is a greater potential pool of bioinformatics *demiurges* (Chapter Seven) waiting to

[101] Dr. Griffiths was at pains to stress that some bioinformaticians could emerge through computer science as well as through the faculty of life sciences.

be trained than ever before. He counter-balances this view however, by stating that there are a large number of biologists who also need a significant amount of training on the computing side.

It would appear then that although the construction of degrees and qualifications can reduce uncertainties by aiding a discipline to stabilise, it also opens up Hood's question further by questioning whether bioinformatics has to emerge from biology, or whether it can also emerge from computer science. If this is the case, then the next generation of bioinformaticians will not learn the computer aspects of bioinformatics through *informal routes* in their 'leisure time'; instead nonbiological scientists, as Collins *et al.* (1998) calls them, may be trained through these more *formal academic routes*.

BIOLOGY AND COMPUTER SCIENCE

Despite the research discovering very few actors who began their training in computer science (they nearly all had a biological background), in the previous account, Dr. Griffiths has already implied that bioinformatics could emerge from within the borders of computer science and Dr. Morris expands on this further:

> "We've certainly been talking here about a Master's in bioinformatics...I suspect its not going to be just from biology and its not going to be just from maths. It's going to be, you know on a good bioinformatics course in University C, you would need people from stats, biology, [and] medical sciences. We have engineering maths as our artificial intelligence [and] we have computer science. All these people would be involved at some level or other in research, and if we are going to do an MSc you would need all these people together to do something. But I have to say, I think we have got a new Professor of artificial intelligence in engineering maths and he is actually starting a course in bioinformatics. So I should think that will be pretty popular."
> **[Dr. Morris: Professor of Statistics]**

In the interview, Dr. Morris explains how the mathematics department in his university has been discussing the idea of developing an MSc course in bioinformatics under the mentorship of a Professor of artificial intelligence. This disclosure begs the question: if courses are being developed in bioinformatics from departments other than biology, could bioinformatics conceivably emerge from computer science? And if we were to take this further, would it be possible

for a computer scientist to make the full transition from a dry lab scientist to a *wet lab* biologist in the same way that *wet lab* biologists are being trained to use computers: "you have got to teach them computers when they know nothing about it" (Dr. Griffiths, p243)?

I believe this disciplinary question needs to be taken into account by the relevant people when creating the frameworks, in this case degree schemes, for normal, stabilised science. Dr. Griffiths has already shown the foresight to mould and educate a more knowledgeable bioinformatician by attempting to set up a professional education scheme to train researchers in bioinformatics. Rather than just relying on self-training and playing around with their home PC, the belief is that a bioinformatics student could learn the knowledges required in both biological and computing disciplines in equal measure. This would be done by creating undergraduate degree schemes in bioinformatics that may promote a process of linguistic socialisation (knowledge-transfer) and hands-on practical training (immersion). Interestingly though, it is those same researchers who learned the skills through self-learning who are promoting the professionalisation of the research area. This begs the further question: How did those researchers learn their own craft, and do they believe the way that they learnt their skills by *doing* rather than *learning* was problematic?

In Chapter Two, I stated that bioinformatics has an extraordinarily wide boundary of interpretation. This is strikingly evident in the NIBDC's official definition of the area. In their report, the committee states that bioinformatics is both the development and application of computational tools of biological data. Using this definition as the starting point, I highlight both the practices of the bioinformatics user (*client*) and the bioinformaticist creator (*developer*) when discussing the role of self-learning in an emerging discipline. I begin with the extract below from Dr. Cherry, who is a user of bioinformatics and is responding to the question of how he learnt to use online databases and run sequences, an activity that falls within the bioinformatics boundary:

> "From my point of view [it was] trial and error. How I became aware is, in the earlier days, I knew for instance Malaria genomes were being

> sequenced and so I dipped into them as regularly as I could. At that point different chromosomes were being sequenced around the world and to look for anything you had to go to about five different sites and do the same search in five different sites. Finally they [were] integrated into one site which made things much easier. But pretty much by trial and error and in some occasions I found genes for colleagues..."
> **[Dr. Cherry: Bioscientist and Molecular Biologist]**

In the example, Dr. Cherry explains how he learnt to navigate and use online warehouses through trial and error. In Chapter Eight I argue that this is a type of self-education that may become more apparent in big biological science. Dr. Cherry has been able to teach himself the skills of applying bioinformatics by browsing the different virtual, sometimes non attributable, Internet sites in search of ubiquitous knowledge. Furthermore, this technique of WWW self-learning has begun to be recognised and integrated by the wider scientific community, with strategic improvements in the information generated by and user friendliness of certain online databases. For example the EBI web-page now provides an online tutorial, 'Toucan', to help educate biological users. A description of this resource, provided by one of the interview respondents, is shown below:

> "Toucan is an educational resource and is designed for beginners in bioinformatics and molecular biology to try and give them an insight into how they can collect data from databases. What they can do with the analysis of the tools [is] to get more information from their sequence data. It also tries to explain why bioinformatics is interesting [and] why there is so much money and resources being put in it, because it is a scary subject for the lay person. Bioinformatics is an all encompassing subject and this gives people an idea of what it is all about and why so much funding is going into it and what can be achieved with it."
> **[Mrs. Eaton: Bioinformatics Educational Resource Leader]**

Through the development of a resource such as 'Toucan', the tacit and/or craft knowledges inherent in creating software packages - which bioinformatician respondents discuss in Chapter Seven - have been turned into a type of generic biological explicit knowledge. This type of knowledge is what Schutz (1943) and Jordan and Lynch (1998) might call "cook-book knowledge"[102]. In this described case, lay people can begin to play around with the tools by following recipe-like

[102] According to Collins and Kusch (1998), software packages are like pocket calculators. They capture the mimeomorphic aspect of actions. Insofar as their output is then slotted into polimorphic action, it is the human that does the rest.

instructions given to them by the online tutorial. The intention behind this resource is to bridge the gap between the self-learning, peripheral individual and professionalised community learning through a partial immersion within the community. Rather than an isolated beginner struggling to learn a web-based technique through trial and error, the tutorial acts as a disembodied, informed knowledge translator, transmitting knowledge to peripheral community actors through a new type of interactive scientific community immersion. Even so, it must be acknowledged at this point that although the Toucan tool exists, very few of the users I interviewed had ever heard about the resource (Chapter Four). Indeed, it was felt that this type of portable packaged knowledge should not replace face-to-face physical interaction (Chapter Seven).

The second example of self-learning is recounted from an interview conducted with Dr. Andrews, a developer of computational tools for biologists. When I asked Dr. Andrews how he got interested in bioinformatics, he explained that it had been a peculiar route:

> "Basically what I was, was originally a zoologist, [and] became a microbiologist. Then in the last three years I have now moved into bioinformatics. It has been a curious route, but basically it has been a conscious decision on my part to move into bioinformatics when I decided that [it] seems to be a productive area and it clearly has a lot of future...
>
> Tthe proceeding three years, I taught myself to programme. Before then I became interested in that area and saw that as a possible future. So I taught myself to programme through text books, incorporated a bit of bioinformatics into the previous contract I was on, wrote this particular project and thankfully it was funded and [so I have] been working in bioinformatics since...But that is how it first started; by self teaching."
> **[Dr. Andrews: Doctor and Lecturer in Biology]**

Dr. Andrews states that he taught himself "through text books" how to programme computers. This admission might add to the work of Collins and Evans (2007) on expertise. In *Rethinking Expertise* they produce a periodic table of expertise - a model that classifies tacit knowledge into either ubiquitous or specialist tacit knowledge. To become a contributory expert in a specialised field, Collins and Evans argue that you have to immerse yourself fully in the activities of that community. Text book knowledge, they state, is at least two levels down

from being a contributory expert: it is either primary source or popular understanding (Collins and Evans 2007, p14). In fact, they argue that text book knowledge is a type of ubiquitous tacit knowledge and do not classify it as specialist tacit knowledge. Dr. Andrews' statement, however, contradicts this assertion. Despite learning knowledge through text books, Dr. Andrews has his own bioinformatics web-page that is home to an extensive array of biological tools used by the biological community. It could be argued therefore that Dr. Andrews is an '*expert*' in bioinformatics. This claim is justified further by reference to the number of actors that use his resource.

Subsequently with the exposition that Dr. Andrews self-learnt to computer programme and recognising Collins and Evans' (2007) work on expertise, the following questions should be discussed: Is it the case that Dr. Andrews has discounted certain interactions with community members when answering my question, or is it the case that experts in new emergent fields do not need as much immersion and understanding of a research field as experts in more stabilised fields? If it is the latter, are cutting-edge fields then characterised by fairly porous boundaries so that interactions are more interdisciplinary than in more settled fields? This might lead to the conclusion that Dr. Andrews has only got, as Collins and Evans (2007) would argue, ubiquitous tacit knowledge in computer science, but coupled with his skills in biology allows him to become a contributory expert in the new field of bioinformatics? If this is true, does this mean that *experts* in new emergent fields are not open to as much critical debate from their peers as those in established fields? Or alternatively, does Collins and Evans' model not consider the development of expertise in new, emerging, cutting-edge fields?

When attempting to discover an answer, the first comment to mention is that there are obviously not as many, if indeed any, so-called established experts in emergent fields in comparison to more mature fields. This means that there are not as many experts for actors to consult with and not as many to immerse themselves into a community. Thus, in the case of Dr. Andrews there would not have been many bioinformaticians to learn from and *follow* since it was researchers like Dr. Andrews who were in fact creating the field; they were as

Ben-David and Collins (1966) call it the *founders* of the field. The second comment to make is that various respondents in the fieldwork acknowledged that they learnt computer programming skills as children, yet no-one suggested they could perform cutting-edge biological experiments at this age. This would suggest that being able to programme, where you get instant feedback telling you that you may have made a mistake, involves a different type of tacit knowledge to performing biological experiments. So if we classify this type of knowledge as ubiquitous tacit knowledge in the way Collins and Evans have, as long as actors have a biological academic background (specialist tacit knowledge), it would appear that it is enough to draw on ubiquitous tacit knowledge to be able to computer programme at a sufficient level to be recognised and identified as a bioinformatician.

Nevertheless, in a period of time when technologies are transforming the nature of research and creating more and more hybrid areas of research (Chapter Five), perhaps it is fair to state that work on expertise must focus on how pioneers of experimental emergent areas become skilled experts. Are their skills refined or are they just clearing the un-trodden path for others to walk? Dr. Griffiths, a pioneer bioinformatician himself, has already stated there is a requirement to create a formal, structured channel of knowledge-transfer (degree schemes). And thus, if they are just clearing paths, then does their value and embodied skill deserve the recognition of an expert, or is expertise only able to be recognised in more established fields of research? Fuller's quote is relevant here:

> "Science may slide into the semantic space of religion and refer more to a set of institutions and rituals than a set of theories and methods. As educational standards fragment, 'knowing' may come to signify special social practices like 'verifying' or certifying, or it may devolve into a casual word, like 'coping' or adapting. Yesterday's oxymorons turn into tomorrow's platitudes" (p141).

Fuller (1997) questions whether science may slip into an institution where *knowing* signifies just *coping* and *adapting* rather than *verifying* and *certifying*. Relating this statement to the case of Dr. Andrews, his role as one of the new breed of multi-skilled scientists is one in which he has been able to adapt his existing talents in order to surface as a '*hybrid chameleon scientist*' on the cusp of

biology and computer science. Trained through formal, academic education to be a biologist, Dr. Andrews has built on and modified his academic skills by becoming a proficient enough computer programmer. Although these techniques are not authenticated by any formal qualifications, the new skills acquired have meant that he can now identify himself (or cope) as both a biologist and a new hybrid bioinformatician: "I was originally a zoologist...[and] moved into bioinformatics" (p247). The admission that it was "a conscious decision" also supports the claim that this choice was a rational calculated decision to move into a niche market and a new flourishing area of genomic science (Chapter Five). In addition to his own self-identification, users of his biological tools are also identifying Dr. Andrews as an expert in bioinformatics by the very act of using his resource. At the conclusion of the interview, Dr. Andrews took pride in stating that his web tools are being used more and more by biological actors. I believe these elucidations reveal one way that scientists claim an identity in the post-genomic era: by adapting their skills and their identity to fit with the trends of the time (Chapter Five).

Another interesting perspective on this conscious decision to pursue a career in new emergent scientific areas is presented by a Ph.D. student in the study. The following quotation is from Mr. Jenkins, a Ph.D. student intimately involved in the field of bioinformatics. When I asked about the need for stabilising educative practices in bioinformatics, his response highlighted the problems in striking a balance between the diversity and specificity of the subjects taught:

> "At the end of the day they were teaching us, so I thought there is this aspect for it, they are obviously covering it. They are speaking to people individually [and so] then obviously there was going to be Masters courses and all sorts. But there again at that point in time, I was doing it in the end, and I ended up doing a Master's in computing because there wasn't a Master's in bioinformatics...But at the end of the day my pure motivation for doing it was getting the computing aspect, because I don't think even now the Master's course they offer are not necessarily relevant for some of the things I want to do because it is so diverse."
> **[Mr. Jenkins: Ph.D. student in Bioinformatics and Mathematical Biology]**

In the last sentence, Mr. Jenkins highlights a dilemma that has been raised throughout this thesis. As already explained, bioinformatics is an amalgamation

of the two autonomous disciplines, computer science and biology, and these disciplines contain large pools of distinct knowledges. This means that when trying to create an MSc course, the course convener may have incredible difficulty in attempting to narrow down the content. This is evident when Mr. Jenkins states: "the course they offer...is so diverse". This diversity is also captured by Delamont, Atkinson and Parry (2000) when commenting on Ph.D. students from different disciplines. They state:

> "The bench scientist's primary concern seems to be 'Can I get my experiment to run?' and the field researcher's concern is 'Can I survive and can I make sense of all this?' The computer scientist's interest is 'Will this programme run?' and 'Will this model yield the right predictions?" (Delamont, Atkinson and Parry 2000, p100).

In the creation of a new course such as an MSc in bioinformatics both a bench scientist's (wet-lab) and a computer scientist's (dry-lab) concerns must be blended. As Mr. Jenkins has explained, this has led to an assortment of contrasting practices, which he finds too diverse for his specific applied tastes. Interestingly, the instability and uncertainty of practices and knowledge is one possible reason why courses in bioinformatics have not filtered down to undergraduate level. Despite Dr. Griffiths declaring that: "we even have a nascent interaction at the undergraduate level where we have the joint degree programme" (p242), this research study has revealed that this is the exception rather than the norm. The problem, once more, is situated on the certainty/uncertainty precipice, which is a key credential when determining if an area has stabilised or not. Delamont *et al.* (2000) suggest that undergraduate level science degrees are laden with convergent and stable knowledge that reinforce scientific knowledge as schematic and definite. However this changes when the student moves up the rung to a doctoral student. They state that:

> "Doctoral science, it transpires, is quite removed from undergraduate experiences where results are predictable and outcomes certain...Ph.D. students find that their experiments go wrong all the time and that successful conclusions, rather than being the outcome of a unitary process, are only achievable through the mutual adjustment of ideas, instruments and activities" (Delamont *et al.* 2000, pp54-55).

The work of Delamont *et al.* (2000) suggests that undergraduate students are shielded from the uncertainties of scientific practice, and it is only at Ph.D. level that they are exposed to dubiety. This then, may be one answer to why bioinformatics has not filtered down to undergraduate level. It seems as if the research area is not yet stable enough to produce a course that will shield undergraduates from the uncertainties inherent in scientific practice (a dilemma that is found in the Arabic proverb at the start of the chapter: "a chameleon does not leave one tree until he is sure of another"). Additionally, this particular take on knowledge creation may also explain another reason why there are not a plethora of postgraduate courses in the area of bioinformatics despite Collins *et al.* (1998) calling for them ten years ago. Doctoral work in cutting-edge science is dependent on what Hacking (1992) characterises as pre-established knowledge. This pre-established knowledge can only be cultivated by pioneers in the early stages of a research area's development. It is then the length of time it takes to generate sufficient pre-established knowledge, allied with unforeseen administrative problems and the institutional logic of universities, that may militate against new disciplines, which determines the time it takes for new research areas to be filtered down through the educative process. The process is also affected by the numerous relationships built up in the different social worlds of science and education. Once these have been socially negotiated and socially validated the outcome may be the emergence of a generation of *followers* to work with the *founders* of the field[103].

THE BIOINFORMATICS CONSULTANT

As discussed in Chapter Seven, a gap has appeared on the boundary between biologists and bioinformatics. This has meant that expertise in bioinformatics may not be acquired through the same communal verification channels that Collins and Evans (2007) describe in more stable areas. To use Hacking (1992) again, one reason behind this is because bioinformatics does not have a cohort of pre-established knowledge to build upon. Moreover, I argue that it also does not have a pre-established process in place to fall back on. The result of this unstable process and lack of established history has meant that knowledge between the

[103] As the research area stabilises it may begin to spawn established experts.

bioinformatician and the biologist is not translated smoothly between the two social worlds (Chapter Seven). The common consensus amongst respondents in the project is that many biologists do not intrinsically understand what bioinformaticians do. Thus I argue, if biology is subsuming bioinformatics, and there is evidence to suggest it is (Bogdanovic and Langlands 2007), it is only embracing the bioinformatics language and the bioinformatics theory, since the actual practices of computer programming and the practices of mathematics are still alien to numerous experienced biologists. This was confirmed recently, when I was questioned by a Professor in a bioscience department, who asked: "How do you understand bioinformatics because I still don't have a clue?" Although, not acknowledging the idea of interactional expertise (Chapters Four and Eight) this did add further evidence to my argument that a gap exists between the two professions. In the extract below Mr. Jenkins further illustrates this knowledge gap:

> "You do have elements of people who do computing as a hobby and have dabbled in it in their work. But what I have found in the Ph.D. so far is that there is this divide, which is half the reason why I feel I would like to fill that divide...You have all these very very good people who are very very good at the biology but don't really have the time to devote themselves to the computer aspect, which is fair enough really, because even the computing aspect is quite consuming...When you have...to train yourself from scratch, then obviously you cannot expect people necessarily who have got established careers in biology or computer science to naturally convert to one or the other without a great investment in time."
> **[Mr. Jenkins: Ph.D. student in Bioinformatics and Mathematical Biology]**

It is revealing that Mr. Jenkins states: 'that even the computing side is quite consuming'. In using the word '*even*', Mr. Jenkins implies that learning the biology is '*even*' more time consuming than learning the computing. This then may give one insight into why many of the bioinformaticians I interviewed had come to the discipline from biology rather than computer science[104]. The fact of the matter is, according to Mr. Jenkins, that they have to invest slightly less time learning the other research area. Nevertheless, as he later states, there is still a significant proportion of time that needs to be invested to learn the computing

[104] However, this could just be the result of the location I performed the research.

side, thus offering another explanation as to why biologists with established careers in biology have not trained themselves in the practices of computer science. Herein lay the reason why Hood (1992) wanted to establish Ph.D. programmes in the area: postgraduate students are fresh, have less time constraints and a greater willingness to learn new skills and practices. This in itself poses the further scenario where, in the future, a generation gap between pre-genome scientists and post-genome scientists may appear. Whether or not there is or will be a generation gap is uncertain, but it is clear from this research that a knowledge gap exists between different professions. In concluding this section I suggest two future scenarios that could help close this chasm and which may influence policy.

The first innovative practice is a change of emphasis in the running of MSc courses and a greater concession of what can actually be achieved in limited time. This is illustrated in a quote by Mr. Jenkins:

> "I think for individuals who have done the experiment and want to get the result; to embrace that is a big undertaking. I don't think it is something that can be undertaken very easily and I think that is why there is a need for people like myself to advise [and] who can take the time and see what people can do and make those suggestions. Maybe the Masters course is more useful as education than training to actually to do it themselves...But maybe the Masters course could be viewed as a way of educating biologists in terms of what you can do and what is available, rather than necessarily leaving them at the end of the day with a set of skills to do it themselves. Because I don't think you can, because you have really got to push it religiously to be able to do it yourself properly. So yes, maybe the masters should be aimed at that, and I think maybe that is what some of them have tried to be and therefore, for those who are very interested and want to do it, it is obviously very misleading for them. I think education in that respect is the only way but then it comes back to this time issue."
> **[Mr. Jenkins: Ph.D. student in Bioinformatics and Mathematical Biology]**

Mr. Jenkins advocates the possibility of changing MSc courses to focus more on the epistemological and ontological questions of bioinformatics, rather than the development of skills. Jenkins believes that the schism that exists now could begin to be closed if MSc courses were run with the intention of educating biologists on the limits and capabilities of bioinformatics, rather than training them in a particular skill set without any theoretical substance. Here, Mr. Jenkins

is suggesting that knowledge and theory (education) is the fundamental basis of a research area. In turn, this would have the potential to create new professions, and it is the development of one of these professions that could lead to a possible second 'chasm closing trajectory'.

The second innovative technique that could help integrate knowledge translation is to create new professions in between the existing boundaries. Essentially, this role would be filled by a type of boundary person referred to in Chapter Eight. In this case, the boundary person would have an applied role and may be called something like a *'bioinformatics consultant'*. Their duty would be to act as a go-between for the two research areas, continually crossing the boundary and facilitating knowledge-transfer. More specifically the *'bioinformatics consultant'* could help biologists come to terms with developments in bioinformatics. Mr. Jenkins discusses this idea:

> "Yes, you could almost see it as a consultancy role in that respect...But maybe there is a market for that, I don't know. It needs a degree of management to show what you need to do. Maybe the best way of going about it is changing their thoughts of what they are trying to achieve from a particular experiment...As I say it is a lot of roles integrated into one, and I think at the moment because you have not got so many people in this area, you end up having to do lots of roles in one person; the doing, the understanding and then the discussion and convincing...But maybe only time will tell with that as more people get involved with it and personally I like the idea of advising people what to do and making suggestions to them rather than sitting there nine to five with all the data sets. I do like the idea of educating people in bioinformatics."
> **[Mr. Jenkins: Ph.D. student in Bioinformatics and Mathematical Biology**

The bioinformatics consultant would utilise the same interactional and contributory techniques illustrated by Dr. Matthews in Chapter Eight to educate bioinformaticians. S/he would bridge the gap between the bioinformatician and the biologist by educating both professions. As Mr. Jenkins suggests, the consultant could educate biologists in bioinformatics, and I argue they could also help translate the biologists' requirements across to the bioinformaticians. The *bioinformatics consultant* would act as a cross-boundary demiurge (Chapters Seven and Eight) and could potentially be a key professional role in aiding the research area stabilise. The role could also aid the filtration of bioinformatics into

mainstream biology and, if successful, could also be the answer to the dilemma of how actors acquire expertise in newly created research areas.

In relating the emergence of new disciplines and innovation in science to the notion of scientific identity however, Ben-David and Collins (1966) state that:

> "...such growth occurs where and when persons become interested in the new idea, not only as intellectual content but also as a potential means of establishing a new intellectual identity and particularly a new occupational role" (p452).

This suggests that bioinformatics may not be fully integrated into higher education teaching until those working in the field (the hybrid chameleon scientists) begin to identify themselves as bioinformaticians. In this regard, bioinformatics course leaders and bioinformatics consultants need to firstly identify themselves as bioinformaticians, which in turn, may impact upon the identities of the *followers*.

CONCLUSION

The creation of intermediate or interstitial knowledge domains and the emergence of interdisciplinarity implies the creation of new types of social actors. These types tend to come in the form of hybrid or *new chameleon* scientists. In this chapter I have illustrated the ways in which pioneering actors are recruited to proteomics and bioinformatics, and some ways that they are acquiring the new skills required to perform post-genomic experiments. Two of these ways have been *experimentation* (trial and error) and the willingness of certain actors to live and teeter on the certainty/uncertainty research precipice. In the research field of bioinformatics this has meant actors attempting to combine the knowledge and skills of both biology and computer science in order to explore uncharted bio-computing territories. Often this has meant combining specialist tacit knowledge in biology with the more ubiquitous tacit knowledge of computer programming. Their expertise is then recognised through a combination of journal article submissions (biological route) and the development of computer programmes (computing route). In proteomics, the chapter has shown how science teachers engage in *two-way mutual linguistic socialisation* and are willing to *share the*

centre stage with students, while in bioinformatics there is a suggestion that a boundary person such as a bioinformatics consultant is required to bridge the gap between education and training.

Thus, it appears that the role of technologies is changing the nature of scientific relationships (Chapter Eight) and scientific expertise. In accordance with this, there is a suggestion that certain actors only need ubiquitous tacit knowledge in some crafts in order to perform as a pioneering expert in a new field (*the founders*). It is also apparent that this may involve only a partial immersion a community or a new type of interactive immersion. Despite the chapter illustrating how *founders* attempt to learn their new craft, it is clear that the construction of academic courses is a stabilising technique. In this regard the chapter has focussed on the processes that are required to be put in place in order for bioinformatics to be integrated in mainstream academic teaching. As such, it has also demonstrated how expertise is constituted differently between pioneers of a discipline (*the founders*) and the proceeding generation (*the followers*) in the sense that the *followers* of a discipline will learn their craft through academic teaching, rather than a combination of academic teaching and learning in their leisure time.

To conclude, I have collated all the different concepts discussed in the chapter to produce a table (Table 9.1.) that illustrates all the emerging identities, expertises and professional roles of interdisciplinary actors working in the fields of proteomics and bioinformatics. I also determine how the actors go about reducing the uncertainties inherent in emerging fields. This is displayed by drawing attention to and classifying some of the techniques used when teaching in a new interdisciplinary area. Furthermore I compare this to how you would imagine, and how I have experienced, these methods in more established fields. It must be stressed though that this is just an exploratory table used to order the ideas discussed in the chapter, and the reason I compare these ideas with more stabilised fields is to highlight the novel practices I found being pioneered in two post-genomic sciences (see p258).

Status	Emerging research area	Stabilised research area
Expert Position	Founders	Followers
Status of Knowledge	Less Certain	More Certain
Expertise	Learnt through a combination of academic teaching and leisure activities	Learnt through academic teaching.
Identity	Hybrid Chameleons (in flux)	Clearly defined intellectual identity
Actors	Interdisciplinary boundary-people	Disciplinary experts
Academic Courses	Hybrid degree courses emanating from different traditional departments	Established academic courses
Immersion	Sometimes partial	Usually Full
Learning	Trial and error	More established techniques
Teaching	Sharing the centre stage	Dominating the centre stage
Language Socialisation	Two-way mutual linguistic socialisation	One-way linguistic socialisation

Table 9.1: Table of Identities, Expertises and Actions Found in Post-Genomic Science.

Table 9.1 summarises the sorts of practices evident in proteomics and bioinformatics. It compares this to more traditional and established fields in order to highlight some of the different and unique procedures involved in the development of a new research field. It demonstrates how, in new areas, knowledge is less certain and professional roles are in flux. The creation of academic roles helps the stabilisation of the field, and the types of practices discussed throughout the chapter are utilised by pioneering founders to aid this process.

CHAPTER TEN:
NEW DISCIPLINES: EMERGENCE AND STABILISATION

INTRODUCTION

This thesis has explored the significant role that standardisation plays in securing coherence for the emergent disciplines of proteomics and bioinformatics. The focal research questions of the ways in which scientists claim an identity in a post-genomic era, how expertise is constituted and how interdisciplinary research is conducted have also been addressed. In this final chapter, I address some of the arguments as they are presented in the thesis, before reflecting on each of the data chapters.

In the early chapters I have demonstrated various ways scientists overcome the uncertainties and ambiguities inherent in scientific practice. Beginning with Shapin and Schaffer (1985), the thesis introduces the reader to an account of modern scientific development. It illustrates how 'scientific' certainties were shaken up with the idea of 'experimental philosophy', and combined with a change in emphasis from relying on what scientists *said* to relying on what scientists *do*, saw the birth of experimental practice. For example, Boyle's inductive approach to producing knowledge meant that, unlike Hobbes who believed in a deductive approach to science, his experimental technique provided limits to the certainties of knowledge. Boyle was more concerned with focussing on what people did agree on and building a sustainable order, rather than concentrating on the problems that uncertainty can provide.

In the same way that Boyle thought that his seventeenth century air pump could bind scientific knowledge and people's understandings together, so the mapping of the Human Genome in the twenty-first century has 'united' contemporary biological knowledge. This can be best illustrated in the way the HGP is talked about as metaphorically writing the book of man (Bodmer and McKie 1995) or

sketching the human blueprint (Shapiro 1992), thus binding biological knowledge together. In fact, it is increasingly difficult to overlook the metaphorical dimension within science today. According to Ahmad (2006): "scientists literally and metaphorically create a world of make-believe through a web of words – some borrowed, some invented, endorsing self belief here and suppressing the beliefs of others there" (p198). Here, Ahmad is writing about how some elementary particles are said to possess flavour or charm and how biological processes reportedly edit, translate or transcribe. Despite the palpable communication and literary benefits of such language, it appears that some metaphors within science are used as rhetorical devices covering-up any knowledge gaps and concealing uncertainties by making complex processes seem more fathomable. Moreover, despite the Human Genome being a symbol of scientific success and a sacred code within the terrain of biology it is not sacrosanct. Ironically, its mapping has brought the limits of scientific knowledge sharply into view by revealing the enormous challenges ahead, while in other quarters some scientists have criticised the holistic science it has engendered. This is best reflected in Chapter Five, where I illustrate the ways in which some reductionist/peripheral scientists have been critical about the amount of funding given to proteomics projects. But this is the point about science, it is not just about the pursuit of truth and the reduction of uncertainties, it is also about how scientists will run their next experiment, from where they receive their next funding, and how they can build up their scientific reputation. This is why this thesis is important. It has explored the origins of new scientific areas of research and examined the intellectual implications of scientific networks, new specialisms and professionalisms, boundary classifications and standardisation techniques.

In the same way Shapin and Schaffer (1985) show how science in the seventeenth century was inherently social, so this thesis has illustrated some of the social orderings in contemporary biological research. Within these new post-genomic fields, similar debates about methodological techniques in the seventeenth century have been recaptured in debates about community standards and emergent practices in the post-genomic stage of omic science. In response, unification of this new type of composition science appears to be driven by a systems biology future where multiple communities will share their practices and understandings,

while still navigating a complex labyrinth of scientific expectations and community regulations. In particular, the thesis has examined how scientists attempt to manage inevitable gaps apparent in new research fields. It would appear then that there is recognition that big biology (Hevly 1992) is a grid (Welsh, Jirotka and Gavaghan 2006), and that an equally functioning social grid must be established to tackle all the organisational problems. Thus, it is fair to state that the global nature of big biology, both geographically and epistemologically, has meant that heterogeneous actors are required to communicate with one another as biological problems become shared matters of concern between actors working on big projects. However, communication and collaboration can lead to homogenisation, which in turn can hide individualism. With fears that the organisational practices of big science will never produce such acclaimed scientific individuals as Einstein, Newton, Faraday or Curie, present day scientific actors may struggle to show their individual value. Consequently, the thesis has examined how scientists claim an identity in the post-genomic era by tracking the emergence and stabilisation of proteomics and bioinformatics (Chapters Five to Nine).

In Chapter Five, standardisation as a from of stabilisation is expounded upon as a question of boundary rearrangement between multiple research areas as a means of attaching themselves to assumed benefits and promises associated with a new era of genomic research. Chapter Six on the other hand concentrates on the establishment of standards as a form of disciplinary identity. It demonstrates how data-reporting standards can inscribe a source of permanency to newly forming research fields, focalising multiple actors to comply with a common recognisable research output. If Chapters Five and Six focussed on issues of proteomics identity, Chapter Seven highlighted the precarious position of bioinformatics in the post-genomic era. Concentrating on the categories of *bioinformatician* and *bioinformaticist* it demonstrates the consequences of being classified as service discipline or a creative specialism in its own right. To this end, the creation of new scientific fields reveals new disciplinary work relationships.

In Chapter Eight I collect data from both the proteomics and bioinformatics world to highlight how interdisciplinary research has become a key trend in omic

science. New standardised ontologies are a fundamental, yet not exclusive, technique to aid inter-disciplinary communication and collaboration. Listing and then analysing these communicative devices Chapter Eight promotes science project managers as types of *matchmakers*, forging collaboration between different fields. Finally, in discussing the emergence and stabilisation of new forms of knowledge, we must not lose sight of the equivalent process whereby new social types or identities are produced and performed. In Chapter Nine, the making of hybrids is discussed in relation to the creation and organisation of new post-genomic degrees.

Here, I draw together the arguments presented in the earlier chapters to demonstrate the social forces present in modern day collaborations.

CHAPTER REFLECTIONS

In Chapter Five I examined the emergence of proteomics by demonstrating how existing fields and technologies attach themselves to this new term. I introduced to the STS literature the idea of a *proto-boundary object* as a concept to describe the level of stabilisation of the research field, and as a kind of pre-cursor to the fully blown boundary object that Star and Griesemer (1989) discuss. Against this background, the chapter argues that scientific research areas need a certain level of stabilisation before they can be construed as fully developed boundary objects. Furthermore, Star (1991) states that a "stabilised network is only stabilised for some" (p43), mainly those who are members of that community of practice or, those who use or maintain it. For those outside that community of practice, the network may still be regarded as embryonic. Subsequently, when evaluating the social classification of a new research area, the chapter analysed the opinions of both actors inside the community (core researchers) and those who are situated more on the peripheries (peripheral researchers). The chapter achieved this by investigating science's relationship with funding and exploring how a scientist's identity is affected by the level of stabilisation of the scientific area.

Continuing to perform research is central to the identity of scientists and the more readily recognisable scientist is rewarded most when they are granted further

funding to continue researching. Packer and Webster (1996) show that even in an era of patent culture located around the commercialization of results, the biggest reward is often to do more research. In regards to this research, the study discovered that actors are willing to follow the buzz-words of the time (to which proteomics is one) if it might mean they can attract further funding for research, but they are not willing to identity themselves as researchers in the new domain until the newly migrated area has stabilised further. In many cases, respondents indicate they feel more comfortable under the traditional stabilised labels since they believe it is more prestigious or acceptable to be regarded as a biologist in academia who is able to perform bioinformatics, rather than a bioinformatician. Much like the flexible standards they create, post-genomic scientists are *malleable* enough to follow new niche markets, but are only confident enough to claim an identity with the more established terms of biologist or chemist. Furthermore, as is discussed in Chapter Seven, one explanation for why researchers identify themselves with the more traditional areas of research rather than the new emerging ones, is because despite the new areas of research being extremely good at attracting research money, the level of maturity and the potential longevity of the two areas are not as transparent. Without any guarantee of how successful the research areas will be, and without any certainty about the shape of their future trajectory, it is less of a risk for actors to identify themselves with their traditional, solidified and mature discipline, rather than the new and potentially only temporary terms of proteomitician or bioinformatician. But even if they do not express a primary disciplinary loyalty to the new field of proteomics, the very existence of pots of research funding attached to the term, attracts and mobilises actors to its boundary.

In a paper by Moore (1965), one of the co-founders of Intel, he describes a fundamental trend in the development of technology that is still prominent today. His 'law' has been interpreted to state that the development of technologies increase exponentially so as to double in power approximately every eighteen to twenty-four months. If we were to apply this 'law' within the setting of proteomics, it is conceivable that in two years time proteomics technology may be twice as powerful as it is today. Apart from illustrating the rapid development of information technology and computational capacity, this 'law' reveals how

technologies and research areas require a stabilisation buffer zone from the period of onset (which I term *phase zero*) to a period of full immersion. However, I argue this is not only a stabilisation period for the technology but also for the actors involved in using that technology. It is during this period from emergence to stabilisation that I argue (if successful) a proto-boundary object may develop into a fully rounded boundary object. The setting in which an original idea of a new research field develops is a complex web and this is often fuelled institutionally. However, stabilisation might only be achieved if the technology and the scientific actors work effectively in order to turn original hype into scientific hope. Chapter Five illustrated the ways in which the stabilisation of a research area would be the successful trajectory of a proto-boundary object developing into a boundary object. In turn this would lead to a reorganisation of scientific practices, research collaborations and material instruments.

The creation of community standards can be instrumental in stabilising knowledge economies. In Chapter Six, I explored the creation of standards in print (data-reporting standards) as best-practice guidelines for the creation of proteomics outputs. The constructed standards help promote international collaboration and assist in identifying and stabilising the research area. Therefore if Chapter Five explored the emanation of proteomics, Chapter Six revealed one way that the field is stabilising.

The creation of a standardised output for a proteomics experiment identifies a research field and helps identify post-genomic scientists' work. Despite the study revealing that terms such as proteomitician or proteomiticist are not yet widely recognised, actors are able to be identified as conducting proteomics work by the appearance of their standardised output. The result of guidelines such as MIAPE mean that proteomics becomes *real*; real in the sense that a group of actors begin producing similar-looking work. With this documentation also comes a source of irreversibility as standards are clearly etched and inscribed into archived history. In this regard the chapter illustrated the ways in which community-based proteomics standards might act as permanent markers celebrating the emergence of a novel research terrain - the standards assist in stabilising the classifications and the classification helps bring order to scientific development. As such

classifications define not merely organisational arrangements, but also an underlying cosmology of knowledge-domains.

The chapter also illustrated how the creation of standards is directed by a particular imagination of the scientific future, that of systems biology. Fujimura (1992) argues that standardised packages are more robust than boundary objects by changing practices on both sides of the boundary. This chapter therefore describes how the *within-communities'* desire to standardise creates a standardised package called proteomics, but that this is directed by a *between-communities'* need to standardise influenced by earlier microarray guidelines such as MIAME. Consequently the data-reporting guidelines can also be seen as standardised packages, scripting new scientific futures into current actions and changing practices on both sides of the boundary.

Despite the virtues of *community-based* standards, Chapter Six also revealed the importance of temporality in their social refinement. As is described in Chapter Five, buzz-words have only a limited temporal hold and, in some cases, the desire to take advantage of this opportune period might be a stronger pull than the value of the standards being created in a more methodical but more mutual manner. During the discussion on 'competing standards' within proteomics we also learnt that the process of standardisation involved two groups –*the standard creators* and *the enforcement agencies*. The relationship between these two groups is instrumental to the success and the trajectory of the standard.

Chapter Seven explored the relationship between the biologist and the bioinformatician through the experiences and opinions of bioinformaticians. To this end it explores the claims and counter claims of various specialisms to their creative scientific contribution with proteomics. Bioinformatics occupies a precarious position here, as either a necessary service or as a creative discipline in its own right. By concentrating on the notion of scientific identity, prestige and credit, the chapter illustrated how finding a niche market that will give actors the opportunity to continue to do work in the future is central to the identity of scientists. To use the terminology of Latour and Woolgar (1979) here, I argued it is *the cycle of credibility* that enables scientists to build identities and careers, and

helps them to continue to conduct research. This supports many of the claims made in Chapter Five about the emergence of proteomics. Chapter Seven illustrated how some '*chameleon bioinformaticians*' developed new computer skills as biologists in preparation for the integration of bioinformatics with biology. It was hoped that this would create cutting-edge post-genomic scientific identities. Despite clear evidence from the actors involved that the *technology* is stabilising, the paradox is that the bioinformaticians argue that their identity is not being fully recognised by biologists or funding agencies (whose committee composition is usually primarily biologists). Rather than their work being acknowledged as knowledge, there is a sense that it is being recognised as information (Lash 2002). As a result, they polarise their own discipline with the classifications *bioinformaticist* and *bioinformatician* founded on the divisive categories of knowledge and information. It is clear that such symbolic classifications around intellectual fields can have implications for the self identities of scientific actors.

Throughout the chapter there is also a suggestion that there are competing identities within the biological field, especially between more traditional reductionist biologists and their more contemporary *omic chameleon* bioinformatician counterparts. This has occurred as a consequence of the introduction of statistics into biology as a way of mass-mapping and assemblage analysing. As Stivers (2001) illustrates, the measure of any scientific technology today is how "visual images and numbers have been replacing language as the primary means of sharing knowledge" (p71). Yet, those numbers may be creating a knowledge gap in the biological world. Chapter Seven endorsed the role of the multi-skilled '*demiurge*', an actor who may turn visual images and numbers into new knowledge. But, the black-boxing of knowledge within automated technology has meant that some of the social processes involved in knowledge-transfer become hidden. Accordingly, the chapter discloses how biologists do not completely understand the inner workings of bioinformatics, and suggests that this lack of transparency leads to a lack of comprehension, which in turn intensifies any divide between the biologist and bioinformatician. This argument could be viewed as the blackbox pessimist view. Nevertheless, as is discussed in Chapter Five, within science there are accounts of failure and accounts of success,

and so Chapter Seven is willing to accept a potential blackbox optimist outlook (Penders *et al.* 2008).

The very notion of interdisciplinarity is predicated on the notion of difference across and within disciplinary fields. Chapter Eight explores the success of interdisciplinary collaborations as mechanisms to overcome the ambiguities, uncertainties and hidden complexities involved in modern day science. In particular, the chapter revealed the importance that space, language and identity play within emergent scientific research areas. The proteomics and bioinformatics settings studied in the research are both physical and virtual spaces as well as being local and dispersed territories. Within these spaces, boundary-people, boundary objects and interdisciplinary networks are essential if knowledge is to be shared and translated smoothly, and if the research area is to stabilise. In the chapter I produce a table (Table 8.1) that emphasises the importance of the both local and global scientific infrastructure. It illustrates the formal/informal, virtual/physical and local/global spaces that exist within omic science. It is within these conducive spaces, which themselves act as boundary objects, where boundary-people (matchmakers and speed-daters) are able to cross disciplinary boundaries. As suggested in Chapters Five and Six, these boundaries may be imagined or real but they all have real organisational and knowledge consequences. Ironically the ever-increasing dependence of computer technology within omic science has also meant that big omic science has the potential to become very small science - the idea of one scientist and their computer. Therefore, Chapter Eight emphasised the significance of face-to-face physical contact in emergent research areas.

Informal spaces or innovative incubators are one of five matchmaking techniques described in Chapter Eight; the others are the *standard language matchmaker*, the *manipulative matchmaker*, *the diplomatic matchmaker*, and *the codebuster*. Over and above this, the chapter also described the social processes and communicative methods that were utilised by boundary-people to overcome uncertainty. Here, the scientist is tested above and beyond their scientific ability as they have to be efficient in communicating, managing, facilitating, ordering, exploring, administrating and translating. It is these techniques that help bridge any

knowledge deficits caused by the migration of heterogeneous knowledge and aids the stabilisation of research activities.

Finally, Beck (2000) asserts that "the inability to know is becoming ever more important in modernity...[and within] highly developed expert rationality" (p217). Therefore, in the absence of specifically trained experts who wield a sense of certainty on matters (the ability to know), the current climate of cutting-edge academic science encourages boundary permeation and regards interdisciplinary, translational research as both a strength and a technique to tame wild terrain. Chapter Eight argued that this is a new type of knowledge production within the academic arena and tentatively coined the term *permodern science*. In permodern science, the relationship between the scientific space and the scientific trailblazers is fundamental for matchmaking knowledge and helping to manoeuvre uncertain pioneering work on to more certain frontiers.

The focus of Chapter Nine was on knowledge-transfer and the constitution of experts in newly developed research areas. The transfer of knowledge from one person or group to another is intrinsically 'social' and highlights how the term *science* is a communal activity. Essentially, if the rudiment of science is to be future orientated (the reason why scientists search for niche markets) then the transfer of scientific knowledge to the next generation of scientists is crucial. Effective knowledge-transfer provides science with a conveyer belt of talented individuals who can build upon and develop the work of the current generation (Chapter Two). Star (1991) claims that: "science is modern in the sense of having a present-orientated outlook, leaving its past for historians" (p14), and so despite the ever integration of statistics and recipe type boxed knowledge (Chapters Seven and Eight), the development and nurturing of skilled 'demiurges' (Chapter Seven) to create, utilise and expand upon that knowledge is fundamental to the functioning of modern science. The process becomes evidently more difficult when the knowledge being created is new and originates from a variety of disciplinary backgrounds. The migration of heterogeneous knowledge, managed by boundary-people using matchmaking techniques (Chapter Eight), needs to be translated to the next generation of scientists. This is done in order to secure the

future of scientific activity, and also to solidify new emergent knowledge; a kind of 'safety in numbers' idea. It is this process which makes uncertain knowledge seem more certain by creating secure foundations for further work - for example the construction of student courses, the filtration of knowledge onto standardised text books and the creation of expert academic positions. The creation of intermediate or interstitial knowledge domains and the emergence of interdisciplinarity in turn implies the creation of new types of social actor. Despite this, Chapter Nine revealed how actors do not currently have an established history of prior knowledge to fall back on and so use the techniques of a *chameleon-type* scientist, migrating between different knowledge sources (and with it changing their identities) in order to validate their experimental work. The creation of professionalised cross-boundary demiurges (some of whom are described in Chapter Eight), may help to bridge the gap and aid the current education of the future teachers of the field, while the creation of academic courses is a further stabilising technique.

CONCLUDING REMARKS

Within the current organisation of scientific practice, it is ironic that science requires uncertainty. Uncertainty permits scope for further progression towards certainty, and the desire to understand things better is the basis for most scientific funding. In this respect, the research fields of proteomics and bioinformatics are also in need of further social science research. For example, it will be interesting to discover to what extent the research areas have stabilised in five years time and in which ways the professional roles of bioinformaticians have been integrated. It will also be fascinating to record whether those working in the field of bioinformatics will remain 'lumped' as bioinformaticians or whether they are 'split' into more specialised forms of work discussed in Chapter Seven. Within the field of STS and specifically within the emerging field of the *Studies of Expertise and Experience* (SEE) (2007), further research is required to explore how someone becomes an *expert* in a new area of research. For example, Collins and Evans (2007) show how you can pass as an expert in a research field through speech (interactional expertise), but I ask, how do you pass as an expert in a field where there are no established experts? Moreover, is interactional expertise a

necessary (but not sufficient) condition for contributory expertise, and if it is, what role would Dr. Matthews occupy on her project if she was not an interactional expert in, for example, mathematics? Could she still be a project leader or a matchmaker? Consequently, the main contribution the study provides the field of STS is to determine how good current STS concepts are for describing events in *emerging* areas of research that have not yet stabilised.

In addition to its contribution on how expertise is constituted, the research also illustrated the role matchmakers play in closing down uncertainties, demonstrated what standards do when implemented, focussed on the *process* of standardisation, and demonstrated how the term *proto-boundary object* is required to describe the initial developments of some boundary objects. The latter of these concepts was to use the actors defined term of buzz-word. For the author, this is one of the peculiarities of the research because apart from my supervisor, I have not met any other social scientist working in the area of proteomics, and have met relatively few working in the bioinformatics setting. In many ways this anomaly is also a direct result of buzz-words and funding, and currently one the main buzz-words and one of the largest pots of funding for STS researchers in the UK is in stem cell research. Nonetheless, it will be interesting, not least since it would support one of the main arguments in the thesis, to note whether proteomics becomes the focus of social scientists' work if it develops and stabilises from a proto-boundary object into a clinical application. For example, during many conference presentations one of the most intriguing questions I have been asked is: What is your view on personalised medicine? To begin with I was puzzled by such questioning since I had never used the terms personalised medicine in presentations and never implied such a trajectory. Latterly, I have realised that questioners were alluding to a potential application trajectory of proteomics. If this were to materialise, I believe that proteomics would gather more interest and, as such, many more social scientists would find themselves working in the field.

The idea of personalised medicine or group-targeted medicine would contain numerous ethical issues requiring further discussion and research. Due to the many intermediate complexities that may mean the research field meandering

down one route or another,[105] I maintain it is currently extremely difficult to predict whether the notion of personalised medicine is a plausible projection. Nevertheless, even tentative debates that connect proteomics to this future illustrate the grand aspirations (if run successfully) of proteomics. The key for scientists is to translate experimental work smoothly from the laboratories to the clinical setting via computer platforms. It is this type of promissory discourse and this imagined future (Brown 2003; Stephens and Lewis 2008; Wainwright *et al.* 2007), which is providing fuel for scientists to re-introduce the seventeenth and eighteenth century principles of modernity with renewed confidence in omic science - 'welcome to the world of point-and-click biology' (Former motto of Incyte Pharmaceutical Company as cited in Penders 2008, p83).

[105] Not least the huge (and unrealistic) amount of funding that would be required to resource such a project.

BIBLIOGRAPHY

Aach, J., Bulyk, M.L., Church, G.M., Comander, J., Derti, A. and Shendure, J. 2001. 'Computational Comparison of Two Draft Sequences of the Human Genome'. *Nature.* 409: pp856-859.

Abbate, J. 2001. *Inventing the Internet.* London: MIT Press.

Abbot, A. 1988. *The System of Professions: An Essay on the Division of Expert Labor.* Chicago, IL: Chicago University Press.

Adam, B., Harris, P. and Lewis, J. 2002. *'Timescape Analysis'. A Foresight Exercise to Help Forward Thinking in Transport and Sectoral Integration.* Inception Report. Vienna: ICCR. pp114-135.

Ahmad, K. 2006. 'Metaphors in the Language of Science'. In Ghotti, M. and Bhatia, V. eds. *New Trends in Specialized Discourse Analysis.* pp197-220.

Akrich, M. 1992. 'The De-scription of Technical Objects'. In Bijker, W. and Law, J. eds. *Shaping Technology/Building Society: Studies in Sociotechnical Change* Cambridge, MA: MIT Press. pp205-224.

Alberts, B.M., Botstein, D., Brenner, S., Cantor, C.R., Doolittle, R.F., Hood, L., McKusick, V.A., Nathans, D., Olson, M.V., Orkin, S., Rosenberg, L.E., Ruddle, F.H., Tilghman, S., Tooze, J. and Watson, J.D. 1988. *'Report of the Committee on Mapping and Sequencing the Human Genome. Board on Basic Biology, Commission of Life Science, National Research Council'.* Washington: National Academy Press.

Aldridge, A. 1993. 'Negotiating Status: Social Scientists and the Anglican Clergy'. *Journal of Contemporary Ethnography.* 22: pp97-112.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. 1990. 'Basic Local Alignment Search Tool'. *Journal of Molecular Biology.* 215: pp409-410.

American Heritage Dictionary of English Language. 4th ed. 2007. Available at: <URL: http://dictionary.reference.com/browse/> [Accessed 11 March 2008].

Amiel, J. 2003. *The Core.* DVD. United Kingdom: Paramount Home Entertainment.

Anderson, N.G., Abajian, V., Anderson, N.L., Johnson, I., McConkey, E., McDermott, W., Neel, j. V., Thomas, S., Whitehead, E.C. 1980. *Report of the Human Protein Index Task* Force'. Rockville: Large Scale Proteomics Group.

Anderson, N.G. and Anderson, N.L. 1982. 'The Human Protein Index'. *Clinical Chemistry.* 28: pp739-748.

Anderson, N.G. and Anderson, N.L. 1998 'Proteome and Proteomics: New Technologies, New Concepts, and New Words'. *Electrophoresis.* 11: pp1853-1861.

Anderson, N.G., Matheson, A. and Anderson, N.L. 2001. 'Back to the Future: The Human Protein Index (HPI) and the Agenda for Post-Proteomic Biology'. *Proteomics.* 1: pp3-12.

Ashton, F.W. 1920. 'Isotopes and Atomic Weights'. *Nature.* 105: pp617.

Atkinson, P.A. 2006 *Everyday Arias: An Operatic Ethnography.* Oxford: Altamira Press.

Atkinson, P.A., Coffey, A., and Delamont, S. 2003. *Key Themes in Qualitative Research: Continuities and Changes.* Walnut Creek: Altamira Press.

Atkinson, P.A. and Glasner, P. 2007. 'Introduction: New Genetic Identities?' In Atkinson, P.A., Glasner, P. and Greenslade, H. eds. *New Genetics, New Identities.* London: Routledge. pp1-10.

Atkinson, P.A. and Hammersley, M. 1994. 'Ethnography and Participant Observation'. In Denzin, N.K. and Lincoln, Y.S. eds. *Handbook of Qualitative Research.* Thousand Oaks, CA: Sage. pp248-261.

Attwood, T.K. and Parry-Smith, D.J. 1999. *Introduction to Bioinformatics.* Harlow: Addison Wesley Longman Limited.

Auffray, C., Imbeaud, S., Roux-Rouquie, M. and Hood, L. 2003. 'From Functional Genomics to Systems Biology: Concepts and Practices'. *C.R. Biologies.* 326: pp879-892.

Baran, P. 1964. 'On Distributed Communications Networks'. *IEEE Transactions and Communications.* 12: pp1-9.

Barnes, B. 1974. *Scientific Knowledge and Sociological Theory.* London: Routledge and Kegan Paul.

Bartlett, A. 2009. *Accomplishing the Sequencing of the Human Genome.* Ph.D. Thesis: Cardiff University.

Beaver, D. 2001. 'Reflections on Scientific Collaboration (and its study): Past, Present and Future'. *Scientometrics.* 52(3): pp365-377.

Beck, U. 1992. *Risk Society: Towards a New Modernity.* London: Sage.

Beck, U. 2000. 'Risk Society Revisited: Theory, Politics and Research Programmes'. In Adam, B., Beck, U. and van Loon, J. eds. *The Risk Society and Beyond: Critical Issues for Social Theory.* pp211-229.

Ben-David, J. and Collins, R. 1966. 'Social Factors in the Origins of New Science: The Case of Psychology'. *American Sociological Review.* 31(4): pp451-465.

Berners-Lee, T. 1989. *Information Management: A Proposal. CERN.* Available at: <URL: http://www.w3.org/(history)1989/proposal.html> [Accessed 8 November 2005].

Beynon, H. 1975. *Working for Ford.* Wakefield: E.P. Publishing.

Bijker, W.E 1995. *Of Bicycles, Bakelites and Bulbs: Towards a Theory of Sociotechnical Change.* London: MIT Press.

Bingen, J., and Busch, L. 2006. *Agricultural Standards: The Shape of the Global Food and Fiber System.* New York: Springer Verlag.

Blackstock, W. and Mann, M. 2001. 'A Boundless Future for Proteomics?' *TRENDS in Biotechnology.* 19(10) (Suppl): ppS1-S2.

Bloom, H. 1985. *Twentieth Century American Literature.* New York: Chelsea House Publishers.

Bloor, D. 1976. *Knowledge and Social Imagery.* London: Routledge and Kegan Paul.

Bodmer, W. and McKie, R. 1995. *The Book of Man: The Human Genome Project and the Quest to Discover Our Genetic Heritage.* New York: Scribner.

Bogdanovic, S. and Langlands, B. 2007. *Bioinformatics: Emerging Tools for Drug Discovery and Development.* London: Business Insights Limited.

Boguski, M.S. 1994. 'Bioinformatics'. *Current Opinion in Genetics and Development.* 4: pp383-388.

Boguski, M.S. and McIntosh, M.W. 2003. 'Biomedical Informatics for Proteomics'. *Nature.* 422: pp233-237.

Bowker, G.C. and Star-Leigh, S. 2000. *Sorting Things Out: Classification and its Consequences.* London: MIT Press.

Boyle, R. 1660. 'New Experiments Physico-Mechanical: Touching the Spring of the Air and their Effects'. *Written by Way of Letter to the Right Honorable Charles Lord Vicount of Dungarvan, Eldest Son to the Earl of Corke.* Oxford: H. Hall.

Brazma, A., Hingamp, P., Quackenbush, J., Sherlock, G., Spellman, P., Stoeckert, C., Aach., J., Ansorge, W., Ball, C.A., Causton, H.C., Gaasterland, T., Glenisson, P., Holstege, F.C., Kim, I.F., Markowitz, V., Matese, J.C., Parkinson, H., Robinson, A., Sarkans, U., Schulze-Kremer, S., Stewart, J., Taylor, R., Vilo, J. and Vingron, M. 2001. 'Minimum Information about a Microarray Experiment

(MIAME)- Toward Standards for Microarray Data'. *Nature Genetics.* 29: pp365-371.

Brenner, S. 1998. 'The Impact of Society on Science'. *Science.* 282(5393): pp1411-1412.

Brewis, I. 2005. *'Proteomics'.* Lecture Notes for Graduate School Bioinformatics Course. 12 July 2005.

British Broadcasting Corporation. 2007. *Scientists Want New Drug Rankings.* Available at: <URL: http://news.bbc.co.uk/1/hi/health/6474053.stm> [Accessed 26 March 2007].

Brooksbank, C., Cameron, G. and Thornton, J. 2005. 'The European Bioinformatics Institute's Data Resources: Towards Systems Biology'. *Nucleic Acids Research.* 33: pp46-53.

Brown, N. 2003. 'Hope Against Hype. Accountability in Biopasts, Presents and Futures'. *Science Studies.* 16: pp 3–21.

Brown, N. and Webster, A. 2004. *New Medical Technologies and Society: Recording Life.* Malden: Polity Press.

Bruun, H. 2007. 'Genomics and the Transformation of Knowledge: The Bioinformatics Challenge'. In Glasner, P., Atkinson, P.A. and Greenslade, H. eds. *New Genetics, New Social Formations.* London: Routledge. pp 187-204

BS-5959. 1980. Specification for Key Numbering System and Layout Charts for Keyboards on Office Machines. *British Standards.* London.

Busch, L. 2008. *Standards and Society: A Research Program.* Cesagen Journal Club Presentation, 1 May 2008. Cardiff.

Callon, M. 1986. 'Some Elements of a Sociology of Translation: Domestication of the Scallops and the Fishermen of St Brieuc Bay.' In Law, J. ed. *Power, Action and Belief: A New Sociology of Knowledge.* London: Routledge and Kegan Paul. pp196-223.

Campbell, A.M. and Heyer, L.J. 2003. *Discovering Genomics, Proteomics and Bioinformatics.* San Francisco, CA: Benjamin Cummings.

Campbell, L. 2004. *Historical Linguistics an Introduction.* Cambridge, MA: MIT Press.

Carr, W. and Kemmis, S. 1986. *Becoming Critical: Education, Knowledge and Action Research.* Lewes: Falmer.

Castells, M. 2000. *The Rise of the Network Society.* 2nd ed. Oxford: Blackwell Publishers.

Clarke, A.E. 1991. 'Social Worlds/Arenas Theory as Organizational Theory'. In Maines, D.R. ed. *Social Organization and Social Processes- Essays in Honour Of Anselm Strauss.* New York: Aldine De Gruyter. pp119-158.

Clarke, A.E and Fujimura, J.H. 1992. 'What Tools? Which Jobs? Why Right?'. In Clarke, A.E. and Fujimura, J.H. eds. *The Right Tools For The Job: At Work in Twentieth-century Life Sciences.* Princeton, NJ: Princeton University Press. pp3-44.

Coenen, A., Marin, H., Park, H., Bakken, S. 2001. 'Collaborative Efforts for Representing Nursing Concepts in Computer-based Systems'. *Journal of the American Medical Informatics Association.* 8: pp202-211.

Coffey, A. and Atkinson, P.A. 1996. *Making Sense of Qualitative Data: Complementary Research Strategies.* London: Sage.

Coffey, A and Atkinson, P.A. 2002. 'Revisiting the Relationship Between Participant Observation and Interviewing', In Gubrium J.F. and Holstein J.A. eds. *Handbook of Interview Research.* Thousand Oaks CA: Sage.

Cohen, J. 2001. 'The Proteomics Payoff'. *Technology Review.* 104: pp55-60.

Collins, F.S. 1995. 'Ahead of Schedule and Under Budget: The Genome Project Passes its Fifth Birthday'. *Proceedings for the National Academy of Sciences.* 92: pp10821-10823.

Collins, F.S., Green, E.D., Guttmacher, A.E. and Guyer, M.S. 2003. 'A Vision for the Future of Genomics Research: A Blueprint for the Genomic Era'. *Nature.* 422: pp835-847.

Collins, F.S., Patrinos, A., Jordan, E., Chakravarti, A., Gesteland, R. and Walters, L. 1998. 'New Goals for the U.S. Human Genome Project: 1998-2003'. *Science.* 282: pp682-689.

Collins, H.M. 1975. 'The Seven Sexes: A Study in the Sociology of a Phenomenon or the Replication of Experiments in Physics'. *Sociology.* 9: pp205-224.

Collins, H.M. 1988. 'Public Experiments and Displays of Virtuosity: The Core-Set Revisited'. *Social Studies of Science.* 18(4): pp725-748.

Collins, H.M. 1992. *Changing Order: Replication and Induction in Scientific Practice.* Chicago, IL: University of Chicago Press.

Collins, H.M. 1999. 'Tantalus and the Aliens: publications, audiences and the search for gravitational waves'. *Social Studies of Science.* 26: pp229-244.

Collins, H.M. 2001. 'Tacit Knowledge, Trust and the Q of Sapphire'. *Social Studies of Science.* 31(1): pp71-85.

Collins, H.M. 2004a. *Gravity's Shadow: The Search for Gravitational Waves.* Chicago IL: University of Chicago Press.

Collins, H.M. 2004b. 'Interactional Expertise as a Third Kind of Knowledge', *Phenomenology and the Cognitive Sciences.* 3(2): pp125-143.

Collins, H.M. 2004c. 'How Do You Know You've Alternated?' *Social Studies of Science.* 34: pp103-106.

Collins, H.M. 2006. *Knowledge and Conferences.* 4S Conference, 4 November 2006. Vancouver.

Collins, H.M. 2007. 'Bicycling on the Moon: Collective Tacit Knowledge and Somatic-limit Tacit Knowledge'. *Organization Studies.* 28(2): pp257-262.

Collins, H. M. 2008. 'Actors' and Analysts' Categories in the Social Analysis of Science.' In Meusburger, P., Welker, M. and Wunder, E. eds. *Clashes of Knowledge.* Dordrecht: Springer. pp101-110.

Collins, H. M. and Evans, R. 2002. 'The Third Wave of Science Studies: Studies of Expertise and Experience'. *Social Studies of Science.* 32(2): pp235-296.

Collins, H.M. and Evans, R. 2007. *Rethinking Expertise.* Chicago, IL: University of Chicago Press.

Collins, H.M., Evans, R., Ribeiro, R. and Hall, M. 2006. 'Experiments with Interactional Expertise'. *Studies in History and Philosophy of Science.* 37(A/4): pp656-674.

Collins, H. M. and Kusch, M. 1998. *The Shape of Actions: What Humans and Machines Can Do.* Cambridge, MA: MIT Press.

Collins, H. M. and Pinch, T. J. 1979. 'The Construction of the Paranormal: Nothing Unscientific is Happening'. In Wallis, R. ed. Sociological Review Monograph. No. 27: On the Margins of Science: The Social Construction of Rejected Knowledge. Keele: Keele University Press. pp237-70.

Constance, L. 1951. 'The Versatile Taxonomist'. *Brittonia.* 7: pp225-231.

Crane, D. 1972. *Invisible Colleges: Diffusion of Knowledge in Scientific Communities.* Chicago, IL: University of Chicago Press.

Cryer, T. 2007. *Research Students' Skills Development Programme September 2007 – July 2008.* Cardiff University.

Daily Mail. 2008. *Outrage as Publicity-Mad Chief Constable says 'Ecstasy is Safer than Aspirin'.* Available at: <URL: http://www.dailymail.co.uk/news/article-505547/Outrage-publicity-mad-chief-constable-says-ecstasy-safer-aspirin.html> [Accessed 5 April 2008].

Darwin, C. 1859. *The Origin of Species.* London: John Murray.

David, P.A. 1992. 'Heroes, Herds and Hysteresis in Technological History'. *Industrial and Corporate Change.* 1(1): pp139.

David, P.A., and Steinmueller, W.E. 1994. 'Economics of Compatibility Standards and Competition in Telecommunication Networks'. *Information Economics and Policy.* 6(3): pp217-241.

Dawkins, R. 1982. *The Extended Phenotype: The Long Reach of the Gene.* Oxford: Oxford University Press.

Dawkins, R. 2007. 'The Irrational Health Service'. *Enemies of Reason.* [Television Programme] Channel Four.

Dayhoff, M.O. 1969. *Atlas of Protein Sequence and Structure.* 4th ed. Silver Springs: National Biomedical Research Foundation.

Dayhoff, M.O. and Eck, R.V. 1970. 'MASSPEC: A Computer Program for Complete Sequence Analysis of Large Proteins from Mass Spectrometry data of a Single Sample'. *Computational Biology Medicine.* 1: pp5-28.

Dayhoff, M.O., Eck, R.V., Chang, M.A., and Sochard, M.R. 1965. *Atlas of Protein Sequence and Structure.* Silver Spring, Maryland: National Biomedical Research Foundation.

Dean, J. 1979. 'Controversy over Classification: A Case Study from the History of Botany'. In Barnes, B. and Shapin, S. eds. *Natural Order.* London: Sage. pp211-230.

Delamont, S., Atkinson, P.A. and Parry, O. 2000. *The Doctoral Experience: Success and Failure in Graduate School.* London: Falmer Press.

Demme, T. 2001. Blow. DVD. Entertainment in Video.

Dennis, C. and Gallagher, R. 2001. eds. *The Human Genome.* Basingstoke: Palgrave.

Denzin, N. 1970. 'Strategies of Multiple Triangulation'. In Denzin, N. ed. *The Research Act in Sociology: A Theoretical Introduction to Sociological Method.* New York: McGraw-Hill. pp297-313.

Denzin, N. and Lincoln, Y.S. 1994. *Handbook of Qualitative Research.* Thousand Oaks, CA: Sage.

Denzin, N. and Lincoln, Y.S. 2003. *The Landscape of Qualitative Research.* 2nd ed. Thousand Oaks, CA: Sage.

Diamond, I and Woodgate, D. 2005. 'Genomics Research in the UK- The Social Science Agenda'. *New Genetics and Society.* 24(2): pp239-252.

Doppelt, G. 1978. 'Kuhn's Epistemological Relativism: An Interpretation and Defense'. *Inquiry.* 21: pp33-86.

Dove, A. 1999. 'Proteomics: Translating Genomics into Products? *Nature Biotechnology.* 17: pp233-236.

Department for Trade and Industry (DTI). 2006. *Science and Technology Funding.* Available at: <URL: http://www.dti.gov.uk/science/science-funding/budget/page28923.html> [Accessed 11 November 2006].

Dulbecco, R. 1986. 'A Turning Point in Cancer Research: Sequencing the Human Genome'. *Science.* 231: pp1055-1056.

Editorial. 2001. 'The Proteome isn't Genome II'. *Nature.* 410 (6830): pp725.
Ellis, E.L. and Delbruck, M. 1939. 'The Growth of Bacteriophage'. *Journal of Genetic Physiology.* 22: pp365-384.

EMBnet. 2006. *What is EMBnet?* Available at: <URL: http://www.embnet.org/node/73> [Accessed 10 January 2006].

Emmerson, R.M., Fretz, R.I. and Shaw, L.L. 2000. 'Participant Observation and Fieldnotes'. In Atkinson, P., Delamont, S., Coffey, A., Lofland, J. and Lofland, L. eds. *Handbook of Ethnography.* London: Sage. pp352-367.

Emmert, D.B., Stoehr, P.J., Stoesser, G. and Cameron, G.N. 1994 'The European Bioinformatics Institute (EBI) Aatabases'. *Nucleic Acids Research.* 22(17): pp3445-3449.

Eriksson, L. 2004. *From Persona to Person: The Unfolding of an (un)Scientific Controversy.* Ph.D. Thesis: Cardiff University.

Eriksson, L. and Webster, A. 2008. 'Standardizing the Unknown: Practicable Pluripotency as Doable Futures'. *Science as Culture.* 17(1): pp57-69.

Etzkowitz, H., Webster, A. and Healy, P. 1998. *Capitalizing Knowledge: New Intersections of Industry and Academia.* New York: SUNY Press.

European Molecular Biology Laboratory – European Bioinformatics Institute (EMBL-EBI). 2005a. 'ENFIN! Computational Systems Biology Comes to Lab Bench Near You'. *EMBL-EBI Press Release.* Hinxton.

European Molecular Biology Laboratory – European Bioinformatics Institute (EMBL-EBI) 2005b. 'EMBL-European Bioinformatics Institute: In a Nutshell.' *EMBL-EBI Press Release*: Hinxton.

Evans, R., Plows, A. and Welsh, I. 2007. 'Towards an Anatomy of Public Engagement with Medical Genetics'. In Atkinson, P.A., Glasner, P. and Greenslade, H. eds. *New Genetics, New Identities.* London: Routledge. pp139-157.

Faletra, P. 2002 'A Note from the Editor: The New Science'. *Journal of Undergraduate Research.* 2: pp1. Available at: <URL: http://educationlink.labworks.org/journal/> [Accessed 10 Januray 2006].

Fara, P. 2003. *Sex, Botany & Empire: The Story of Carl Linnaeus and Joseph Banks*. Cambridge: Icon Books.

Fenn, J.B., Mann, M., Meng, C.K., Wong, S.F. and Whitehouse, C.M. 1989. 'Electrospray Ionization for Mass Spectrometry of Large Biomolecules'. *Science.* 246: pp64-71.

Fenstermacher, D. 2004. 'Introduction to Bioinformatics'. *Journal of the American Society for Information Science and Technology.* 56(5): pp440-446.

Feyerabend, P. 1993. *Against Method.* 3rd ed. London: Verso.

Fielding, N. 1988. *Joining Forces: Police Training, Socialization and Occupational Competence.* London: Routledge.

Fields, S. 2001. 'Proteomics in Genomeland'. *Science.* 291(5507): pp1221-1224.

Fine, G.A. 2006. 'Ground Truth: Verification Games in Operational Meteorology'. *Journal of Contemporary Ethnography.* 35(1): pp3-23.

Fleck, L. 1979. *Genesis and the Development of Scientific Fact.* Chicago, IL: University of Chicago Press.

Flick, U. 1998. *An Introduction to Qualitative Research.* London: Sage.

Flick, U. 2002. *An Introduction to Qualitative Research.* 2nd ed. London: Sage.

Frank, A. 1995. *The Wounded Storyteller: Body Illness and Ethics.* Chicago IL: University of Chicago Press.

Frauenfelder, H. 2002. 'Proteins: Paradigms of Complexity'. *Proceedings of the National Academy of Science.* 99(1): pp2479-2480.

Fujimura, J.H. 1988. 'The Molecular Biological Bandwagon in Cancer Research: Where Social Worlds Meet' *Social Problems.* 35: pp261-283.

Fujimura, J.H. 1992. 'Crafting Science: Standardized Packages, Boundary Objects, and Translation'. In Pickering, A. ed. *Science as Practice and Culture.* Chicago, IL: University of Chicago Press. pp168-211.

Fujimura, J.H. 1996. *Crafting Science: A Sociohistory of the Quest for the Genetics of Cancer.* Cambridge, MA: Harvard University Press.

Fujimura, J.H. 2005. 'Postgenomic Futures: Translations Across the Machine-Nature Border in Systems Biology'. *New Genetics and Society.* 24(2): pp195-225.

Fuller, S. 1997. *Science.* Buckingham: Open University Press.

Functowicz, S.O. and Ravetz, J. 1993. 'Science for the Post-Normal Age.' *Futures* 25: pp739–55.

Galison, P.L. 1997. *Image and Logic: A Material Culture of Microphysics.* Chicago, IL: University of Chicago Press.

Garavelli, J. 2002. 'The Proteome Challenge'. *Minutes of the 2140th Meeting.* Philosophical Society of Washington.

Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R.D., Bairoch, A. 2003. 'ExPASy: The Proteomics Server for In-depth Protein Knowledge and Analysis'. *Nucleic Acids Research.* 31(13): pp3784-3788.

Geesink, I. 2006. *Risk Regulation of Tissue Engineering in the EU: A Political Economy of Medicine.* Ph.D. Thesis: Cardiff University.

Geesink, I., Prainsack, B., Franklin, S. 2008. 'Guest Editorial: Stem Cell Stories 1998-2008'. *Special Issue Science as Culture.* 17(1): pp1-11.

Gibbons, M. 2007. *How to Think about Science* [WWW Radio Programme]. Available at: URL: <http://www.cbc.ca/ideas/features/science/index.html#episode20> [Accessed 25 June 2008].

Gibbons, M., Limoges, C., Nowotny, H., Schwartzman, S., Scott, P. and Trow, M. 1994. *The New Production of Knowledge.* London: Sage.

Giddens, A. 1990. *The Consequences of Modernity.* Cambridge: Polity Press.

Giddens, A. 1991. *Modernity and Self-Indentity. Self and Society in the Late Modern Age.* Cambridge: Polity Press.

Gieryn, T.F. 1983. 'Boundary-Work and the Demarcation of Science from Non-Science: Strains and Interests in Professional Ideologies of Scientists'. *American Sociological Review.* 48: pp781-795.

Gieryn, T.F. 1999. *Cultural Boundaries of Science: Credibility on the Line.* Chicago, IL: University of Chicago Press.

Gilbert, N. and Mulkay, M. 1984. *Opening Pandora's Box: A Sociological Analysis of Scientist's Discourse.* Cambridge: Cambridge University Press.

Gohlke, R. 1959. 'Time-of-Flight Mass Spectrometry and Gas Liquid Partition Chromatography'. *Analytical Chemistry.* 31: pp535-541.

Gorman, M. 2002. 'Levels of Expertise and Trading Zones'. *Social Studies of Science.* 32(6): pp933-938.

Gould, S.J. 1989. *Wonderful Life.* Harmondsworth: Penguin.

Gould, S.J. 1991. *Bully for Brontosaurus.* New York: W.W.Norton and Company.

Granovetter, M.S. 1973. 'The Strength of Weak Ties'. *American Journal of Sociology.* 78(6): pp1360-1380.

Granovetter, M.S. 1983. 'The Strength of Weak Ties: A Network Theory Revisited'. *Sociological Theory.* 1: pp201-233.

Graves, R. 1955. *The Greek Myths.* London: Penguin.

Griffiths, R.T. 1949. 'The Minimotion Typewriter Keyboard'. *Journal of Franklin Institute.* 248: pp399-436.

Gross, M. 1998. 'Protein Folding: Think Globally, (Inter)act Locally'. *Current Biology.* 8(9): R308-9.

Gunnarsdottir, K. 2005. 'Scientific Journal Publications: On the Role of Electronic Preprint Exchange in the Distribution of Scientific Literature'. *Social Studies of Science.* 35(4): pp549-580.

Gupta, P. and Guglani, L. 2001. 'Proteomics: Challenge for the New Millennium'. *Indian Pediatrics.* 38: pp875-883.

Hacking, I. 1992. 'The Self-Vindication of the Laboratory Sciences'. In Pickering, A. ed. *Science as Practice and Culture.* Chicago, IL: Chicago University Press. pp29-64.

Hammersley, M. 2003. 'Analytics are No Substitute for Methodology'. *Sociology.* 37(2): pp339-351.

Hammersley, M. and Atkinson, P.A. 1995. *Ethnography: Principles in Practice.* 2nd ed. London: Routledge.

Hanash, S. 2004. 'HUPO Initiatives Relevant to Clinical Proteomics'. *Molecular and Cellular Proteomics.* 3(4): pp298-301.

Hanseth, O., Monteiro, E., Hatling, M. 1996. 'Developing Information Infrastructure: The Tension Between Standardization and Flexibility'. *Science, Technology and Human Values.* 21(4): pp407-426.

Harwood, R. 2002. *Biochemistry.* Cambridge: Cambridge University Press.

Hedgecoe, A. 2001. 'Schizophrenia and the Narrative of Enlightened Geneticization'. *Social Studies of Science.* 31(6): pp.875-911.

Hedgecoe, A. 2003. 'Terminology and the Construction of Scientific Disciplines: The Case of Pharmacogenomics'. *Science, Technology and Human Values.* 28(4): pp513-37.

Henzel, W.J., Billeci, T.M., Stults, J.T., Wong, S.C., Grimley, C., & Watanabe, C. 1993. 'Identifying Proteins from 2-Dimensional Gels by Molecular Mass Searching of Peptide Fragments in Protein Sequence Databases'. *Proceedings of the National Academy of Sciences of the United States of America.* 90(11): pp5011-5015.

Hermjakob, H., Taylor, C., Zhu, W. and Apweiler, R. 2004. *The HUPO Proteomics Standard Initiative.* (HUPO-PSI). Available at: <URL: http://www.icsb.org/ismb2004/posters/hheATebi.ac.uk_326.html> [Accessed 20 July 2005].

Heslop-Harrison, J. 1953. *New Concepts in Flowering Plant Taxonomy.* London: Heinemann.

Hevly, B.W. 1992. 'Reflections on Big Science and Big History'. In Galison, P.L. and Hevly, B.W. eds. *Big Science: The Growth of Large-scale* Research. Palo Alto, CA: Stanford University Press. pp355-363.

Hillenkamp, F., Karas, M. Beavis, R.C. and Chait, B.T. 1991. 'Matrix-assisted Laser Desorption/Ionization Mass Spectrometry of Biopolymers.' *Analytical Chemistry* 63(24): pp1193A-1203A.

Hine, C. 2006. eds. *New Infrastructure for Knowledge Production: Understanding E-Science.* London: Information Science Publishing.

Hirs, C.H.W., Moore, S. and Stein, W.H. 1960. 'The Sequence of Amino Acid Residues in Performic Acid-oxidized Ribonuclease'. *Journal of Biological Chemistry.* 235: pp633-647.

Hood, L. 1992. 'Biology and Medicine in the Twenty-First Century'. In Kevles, D.J. and Hood, L. eds. *The Code of Codes: Scientific and Social Issues in the Human Genome Project.* Cambridge, MA: Harvard University Press. pp136-163.

Hood, L. 2003. 'Systems Biology: Integrating Technology, Biology and Computation'. *Mechanisms of Ageing and Development.* 124: pp9-16.

Huerta, M., Haseltine, F., Liu, Y., Downing, G. and Seto, B. 2000. *NIH Working Definition of Bioinformatics and Computational Biology.* Available at: <URL: http://www.bisti.nih.gov/docs/CompuBioDef.pdf> [Accessed 10 January 2006].

HUPO-PSI. 2007. *Home Web Page.* Available at: <URL: http://www.psidev.info/> [Accessed 3 March 2005].

HUPOST. 2005. 'Inauguration of HUPO International Headquarters'. *The Newsletter of the Human Proteome Organisation.* 1(1): pp1-6.

HUPOST. 2006. 'Special Feature on the HUPO Educational and Training Initiative'. *The Newsletter of the Human Proteome Organisation.* 1(2): pp1-6.

HUPOST. 2007. 'Fostering International Proteomic Initiatives to Better Understand Human Disease'. *The Newsletter of the Human Proteome Organisation.* 2(1): pp1-6.

Human Genome Organisation. 2005. *Home Web Page.* Available at: <URL: http://www.hugo-international.org> [Accessed 3 December 2005].

Human Proteome Organisation. 2005. *Home Web Page* Available at: <URL: www.hupo.org> [Accessed 14 February 2005].

Human Proteome Organisation. 2006. *Home Web Page. Available at*: <URL: www.hupo.org> [Accessed 8 November 2006].

Ideker, T., Galitski, T. and Hood, L. 2001. 'A New Approach to Decoding Life: Systems Biology'. *Annual Review of Genomics and Human Genetics.* 2: pp343-372.

ISO-4169. 1979. Standards for Office Machines, Keyboards, Key Numbering System and Layout Charts. *International Organisation for Standardization.* Geneva.

ISO-9241. 1998. Ergonomic Requirements for Office Work with Visual Display Terminals (VDTs). *International Organisation for Standardization.* Geneva.

Israel, J. 2006. *Enlightenment Contested: Philosophy, Modernity and the Emancipation of Man, 1670-1752.* Oxford: Oxford University Press.

Jacobson, R. 1959. 'On Linguistic Aspects of Translation'. In Brower, R.A. ed. *On Translation* Cambridge, MA: Harvard University Press. pp232-239.

Jasanoff, S. 2003. 'Breaking the Waves in Science Studies: Comment on H.M. Collins and Robert Evans, 'The Third Wave of Science Studies". *Social Studies of Science.* 33(3): pp389-400.

Jasanoff, S. ed. 2004. *States of Knowledge: The Co-production of Science and Social Order.* London: Routledge.

Jenkins, A.C. 1978. *The Naturalists: Pioneers of Natural History.* New York: Mayflower Books.

Jenkins, L. 2007. 'Bycatch: Interactional Expertise, Dolphins and the U.S. Tuna Fishery'. *Studies in the History and Philosophy of Science Part A.* 38(4): pp698-712.

Jordan, K. and Lynch, M. 1998. 'The Dissemination, Standardization and Routinazation of a Molecular Biological Technique'. *Social Studies of Science.* 28 (5/6): pp773-800.

Kadushin, C. 1966. 'The Friends and Supporters of Psychotherapy: On Social Circle in Urban Life'. *American Sociological Review*. 31: pp786-802.

Kadushin, C. 1968. 'Power, Influence and Social Circles: A New Methodology for Studying Opinion Makers'. *American Sociological Review*. 33: 685-699.

Kaiser, J. 2002. 'Public-Private Group Maps out Initiatives'. *Science*. 296(5569): pp827-828.

Kay, L.E. 2000. *Who Wrote the Book of Life? A History of the Genetic Code.* Stanford: Stanford University Press.

Keating, P., Limoges, C. and Cambrosio, A. 1999. 'The Automated Laboratory: the Generation and Replication of Work in Molecular Genetics'. In Mendelsohn, E. and Fortun, M. eds. *The Practice of Human Genetics, Sociology of Sciences Yearbook 19.* Dordrecht: Kluwer. pp125-142.

Kellenberger, E. 2004. 'The Evolution of Molecular Biology'. *EMBO Reports.* 5(6): pp546-549.

Keller, E.F. 2000. *The Century of the Gene.* Cambridge, MA: Harvard University Press.

Kersey, P., Bower, L., Morris, L., Horne, A., Petryszak, R., Kanz, C., Kanapin, A., Das, U., Michoud, K., Phan, I., Gattiker, A., Kulikova, T., Faruque, N., Duggan, K., Mclaren, P., Reimholz, B., Duret, L., Penel, S., Reuter, I. and Apweiler, R. 2005. 'Integr8 and Genome Reviews: Integrated Views of Complete Genomes and Proteomes'. *Nucleic Acids*. 33: pp297-302.

Kevles, D.J. and Hood, L. 1992. eds. *The Code of Codes: Scientific and Social issues in The Human Genome Project.* Cambridge, MA: Harvard University Press.

Klose, J. and Spielmann, H. 1975. 'Gel Isoelectric Focusing of Mouse Lactate Dehydydrogenase: Heterogeneity of the Isoenzymes A4 and X4' *Biochemical Genetics.* 9-10: pp707-720.

Knight, K. 1997. 'Automating Knowledge Acquisition For Machine Translation – Natural Language Processing'. *AI Magazine*. 18(4): pp81-96.

Knorr-Cetina, K.D. 1981. *The Manufacture of Knowledge: An Essay on the Constructivist and Contextual Nature of Science*. Oxford: Pergamon Press.

Knorr-Cetina, K.D. 1983. 'The Ethnographic Study of Scientific Work: Towards a Constructivist Interpretation of Science'. In Knorr-Cetina, K.D. and Mulkay, M. *Science Observed: Perspectives on the Social Study of Science*. London: Sage. pp115-140.

Knorr-Cetina, K.D. 1999. *Epistemic Cultures: How the Sciences Make Knowledge.* London: Harvard University Press.

Korf, I., Yandell, M., Bedell, J. 2004. *Blast.* Sebastopol, CA: O'Reilly.

Krefting, N. 2005. 'Biological Classification Schemes and their Relation to Societies' Available at: <URL:http://serendip.brynmawr.edu/biology/b103/f05/web1/nkrefting.html> [Accessed 8 January 2006].

Kuhn, T. 1962. *The Structure of Scientific Revolutions.* 1st ed. Chicago, IL: University of Chicago Press.

Kuhn, T. 1970. *The Structure of Scientific Revolutions.* 2nd ed. Chicago, IL: University of Chicago Press.

Kuhn, T. 1996. *The Structure of Scientific Revolutions.* 3rd ed. London: University of Chicago Press.

Lander, E.S., Linton, L.M. Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., *et al.* 2001. 'Initial Sequencing and Analysis of the Human Genome'. *Nature.* 409: pp860-921.

Lash, S. 2002. *Critique of Information.* London: Sage.

Latour, B. 1992. *We Have Never Been Modern.* Cambridge, MA: Harvard University Press.

Latour, B. and Woolgar, S. 1979. *Laboratory life: the Construction of Scientific Facts.* Princeton, NJ: Princeton University Press.

Law, J. 1973. 'The Development of Specialities in Science: The Case of X-ray Protein Crystallography'. *Science Studies.* 3(3): pp275-303.

Law, J. 1994. *Organising Modernity.* Oxford: Blackwell Publishers.

Lemaine, G., Macleod, R., Mulkay, M. and Weingart, P. 1976. 'Problems in the Emergence of New Disciplines'. In G. Lemaine, R. Macleod, M. Mulkay and P. Weingart (eds) *Perspectives on the Emergence of Scientific Disciplines.* Chicago, IL: Mouton and Co. pp1-23.

Lesk, A.M. 2002. *Introduction to Bioinformatics.* Oxford: Oxford University Press.

Lewontin, R. 1992. *Biology as Ideology: The Doctrine of DNA.* New York: HarperPerennial.

Lewis, J. 2006. 'Making Order out of a Contested Disorder: The Utilisation of Online Support Groups in Social Science'. *Qualitative Researcher.* 3: pp4-7.

Liebler, D.C. 2002. *Introduction to Proteomics: Tools for the New Biology.* New Jersey: Humana Press.

Liew, C., Ma, J., Tang, H., Zheng, R. and Dempsey, A. 2006. 'The Peripheral Blood Transcriptome Dynamically Reflects System Wide Biology: Potential Diagnostic Tool'. *Journal of Laboratory and Clinical Medicine.* 147(3): pp126-132.

Lingard, L., Schryer, C.F., Spafford, M.M. and Campbell, S.L. 2007. 'Negotiating the Politics of Identity in an Interdisciplinary Research Team'. *Qualitative Research.* 7: pp501-519.

Linnaeus, C 1758. *Systema Naturae Per Regna Tria Naturae.* 10th ed. Stockholm: Holmaie.

Liotta, L.A. and Petricoin, E.F. 2000. 'Beyond the Genome to Tissue Proteomics'. *Breast Cancer Research.* 2: pp13-14.

LivingInternet. 2007. *Home Web Page.* Available at: <URL: http://www.livinginternet.com> [Accessed 3 March 2007].

Livingstone, D.N. 2002. *Science, Space and Hermeneutics.* Hettner-Lecture 2001: Department of Geography, University of Heidelberg.

Lynch, M. 2001. 'The Contingencies of Social Constructionism'. *Economy and Society.* 30(2): pp240-254.

MacBeath, G. 2002. 'Protein Microarrays and Proteomics'. *Nature Genetics.* 32: pp526-532.

MacKenzie, A. 1996. 'Undecidability: The History and Time of the Universal Turing Machine'. *Configurations* 4(3): pp359-379.

MacKenzie, A. 2005. 'The Performativity of Code: Software and Cultures of Circulation'. *Theory, Culture and Society.* 22(1): pp71-92.

MacKenzie, D. 1990. *Inventing Accuracy: A Historical Sociology of Nuclear Missile Guidance.* Cambridge, MA: MIT Press.

Marrett, C.B. 1987. 'Closure and Controversy: Three Mile Island'. In Engelhardt Jr H.T. and Caplan, A.L. eds. *Scientific Controversies: Case Studies in the Resolution and Closure of Disputes in Science Technology.* Cambridge: Cambridge University Press.

Martens, L., Hermjakob, H., Jones, P., Taylor, C., Gervaet, K., Vandekerckhove, J. and Apweiler, R. 2005. 'PRIDE: the PRoteomics IDEntifications Database'. *Proteomics.* 5 (13): pp3537-3545.

Marx, K [1859] (1999) A Contribution to the Critique of Political Economy' Available at: URL <http://www.marxists.org/archive/marx/works/1859/critique-pol-economy/> [Accessed 30 June 2008].

Mason, J. 2006. 'Mixing Methods in a Qualitatively Driven Way'. *Qualitative Research.* 6: pp9-25.

Maxam, A.M. and Gilbert, W.A. 1977. 'A New Method for Sequencing DNA'. *Proceedings of the National Academy of Science.* 74: pp550-564.

May, T. 2005. *Social Research: Issues, Methods and Process.* 3rd ed. Berkshire: Open University.

McArthur, T. 1999. 'On the Origin and Nature of Standard English'. *World Englishes.* 18(2): pp161-169.

McIntyre, S.F. 2005. 'From Proteins to Systems: The British Society for Proteomics Research (BSPR) meeting'. *Proteomics.* 5: pp3828-3830.

McKusick, V.A. 1969. On Lumpers and Splitters, or the Nosology of Genetic Disease. Perspectives in Biological Medicine. Winter; 12(2): 298-312.

McKusick, V.A. 1989. 'The Human Genome Organisation History, Purpose and Membership'. *Genomics.* (5): pp385.

McKusick, V.A. 1997. 'Genomics: Structural and Functional Studies of Genomes'. *Genomics.* 45(2): pp 244-249.

McNally, R. 2005. 'Sociomics! Using the IssueCrawler to Map, Monitor and Engage with the Global Proteomics Research Network'. *Proteomics.* 5(12): pp1-7.

McNally, R. 2008. 'Sociomics: CESAGen Multidisciplinary Workshop on the Transformation of Knowledge Production in the Biosciences, and its Consequences'. *Proteomics.* 8: pp222-224.

McNally, R. and Glasner, P. 2005. 'Sociomics: Locating and Analysing Proteomics Networks on the World Wide Web'. *Molecular and Cellular Proteomics.* 4(8): ppS384.

McNally, R. and Glasner, P. 2007. 'Survival of the Gene: 21st Century Visions from Genomics, Proteomics and Systems Biology'. In Glasner, P. and Atkinson, P. eds. *New Genetics, New Social Formations.* London: Routledge. pp253-278.

McNally, R. and Glasner, P. and Wynne 2007. *'Sociomics': A Multidisciplinary Workshop on the Transformation of Knowledge Production in the Biosciences, and its Consequences.* Programme Notes for ESRC Cesagen Dissemination Event. 23-24 July. Hinxton

Mehnert, R. and Cravedi, K. 2005. 'Public Collections of DNA and RNA Sequence Reach 100 Gigabases'. *National Institutes of Health Press Release.* August 22, 2005.

Mennella, T. 2003. 'The Out-of-Hand Omnipresent Ome'. *The Scientist.* 17(12): pp52.

Merrick, A. 2003. 'The Human Proteome Organisation and Environmental Health – HUPO – Commentary'. *Environmental Health Perspectives.* 111(6): pp797-801.

Merton, R. 1943. *The Sociology of Science.* Chicago, IL: University of Chicago Press.

Merton, R.K. 1970. *Science, Technology and Society in Seventeenth Century England.* New York: Howard Fertig.

Mesarovic, M.D. 1968. 'Systems Theory and Biology – View of a Theoretician'. In Mesarovic, M.D. ed. *System Theory and Biology.* New York: Springer Verlag. pp 58-87.

Miller, C.J. and Attwood, T. 2003. 'Bioinformatics Goes Back to the Future'. *Nature Reviews: Molecular Cell Biology.* 4: pp157-162.

Mills, C.W. 1956. *The Power Elite.* New York: Oxford University Press.

Misa, T.J. 1992. 'Controversy and Closure in Technological Change: Constructing *Steel*'. In Bijker, W.E. and Law, J. eds. *Shaping Technology? Building Society: Studies in Sociotechnical Change.* London: MIT Press. pp109-139.

Moody, G. 2004. *Digital Code of Life: How Bioinformatics is Revolutionizing Science, Medicine and Business.* New Jersey: John Wiley and Sons:

Moore, G.E. 1965. 'Cramming More Components onto Integrated Circuits'. *Electronics.* 38(8): pp1-4.

Morange, M. 2006. 'Post-Genomics, Between Reduction and Emergence' *Synthese.* 151(3): pp355-360.

Mulkay, M.J. 1976. 'Norms and Ideology'. *Social Science Information.* 15: pp637-656.

Mulkay, M.J. 1979. 'Science and the Sociology of Knowledge'. In Bottomore, T.B. and Mulkay, M.J. eds. *Controversies in Sociology 8.* London: Allen and Unwin. pp1-122.

Mulkay, M.J. and Gilbert, N. 1984. *Opening Pandora's Box: A Sociological Analysis of Scientists' Discourse.* Cambridge: Cambridge University Press.

Nadel, S.F. 1956. 'The Concept of Social Elites'. *International Social Science Bulletin.* 8: pp413-424.

National Center for Biotechnology Information (NCBI). 2005. *Home Web Page.* Available at: URL <http://www.ncbi.nlm.nih.gov/> [Accessed 22 October 2005].

Nature Editorial. 1999. 'The Promise of Proteomics'. *Nature.* 402(6763): pp703.

Nowotny, H. 2008. *'The Potential of Transdisciplianirity'.* Available at: <URL: http://www.interdisciplines.org/interdisciplinarity/papers/5/language/en> [Accessed 11 August 2008].

Noyes, J. 1998. 'QWERTY- the Immortal Keyboard'. *Computing and Control Engineering Journal.* pp117-122.

O'Connell, J. 1993. 'Meterology: the Creation of Universality by the Circulation of Particulars'. *Social Studies of Science.* 23: pp129-174.

Odendahl, T. and Shaw, A.M. 2002. 'Interviewing Elites'. In Gubrium, J.F. and Holstein, J.A. eds. *Handbook of Interview Research: Context and Method.* Thousand Oaks, CA: Sage.

O'Farrell, P.H. 1975. 'High Resolution Two-Dimensional Electrophoresis of Proteins'. *Journal of Biological Chemistry.* 250(10): pp4007-4021.

Office of the European Union. 2005. Available at: <URL: http://oami.eu.int/> [Accessed August 24th 2005].

Olson, M.V. 1993. 'The Human Genome'. *Proceedings of the National Academy of Sciences.* 90: pp4338-4344.

Orchard, S., Hermjakob, H. and Apweiler, R. 2003. 'The Proteomics Standards Initiative'. *Proteomics.* 3: pp1374-1376.

Orchard, S., Hermjakob, H. and Apweiler, R. 2004. 'Proteomics and Data Standardisation'. *Drug Discovery Today.* BIOSILICO 2: pp91-93.

Orchard, S., Hermjakob, H., Taylor, C.F., Potthast, F., Jones, P., Zhu, W., Julian, R.K. and Apweiler, R. 2005a. 'Further Steps in Standardisation. Report of the Second Annual Proteomics Standards Initiative Spring workshop (Siena, Italy 17th-20th April 2005)'. *Proteomics.* 5(14): pp3552-3555.

Orchard, S., Montecchi-Pallazi, L., Hermjakob, H. and Apweiler, R. 2005b. 'The Use of Common Ontologies and Controlled Vocabularies to Enable Data Exchange Deposition for Complex Proteomic Experiments'. *Pacific Symposium on Biocomputing.* 10: pp186-196.

Owen, J. 2007. 'Cannabis an Apology'. *Independent on Sunday.* 18 March 2007. Available at: <URL: http://www.independent.co.uk/life-style/health-and-

wellbeing/health-news/cannabis-an-apology-440730.html> [Accessed 9 August 2007].

Packer, K. and Webster, A. 1996. 'Patenting Culture in Science: Reinventing the Scientific Wheel of Credibility', *Science. Technology and Human Values.* 21: pp427-453.

Pandey, A. and Mann, M. 2000. 'Proteomics to Study Genes and Genomes' *Nature.* 405: pp837-846.

Patterson, S.D. and Aebersold, R.H. 2003. 'Proteomics: the first decade and beyond'. *Nature Genetics Supplement.* 33: p311-323.

Penders, B. 2008 'From Seeking Health to Finding Healths: The Politics of Large-Scale Cooperation in Nutrition Science'. Maatricht: Universitaire Pers Maastricht.

Penders, B., Horstman, K, Vos, R. 2008. 'Walking the Line Between Lab and Computation: The Moist Zone'. *Bioscience.* 58(8): pp747-755.

Persidis, A. 1999. 'Bioinformatics'. *Nature Biotechnology.* 17: pp828-830.

Phillips, E.M., Pugh, D.S. 2005. *How to Get a PhD.* 4th ed. Maidenhead: Open University Press.

Pickering, A.R. 1984. *Constructing Quarks: A Sociological History of Particle Physics.* Edinburgh: Edinburgh University Press.

Pickering, A.R. 1992. 'From Science as Knowledge to Science as Practice'. In Pickering, A.R. ed. *Science as Practice and Culture.* Chicago, IL: University of Chicago Press.

Pinch, T. and Trocco, F. 2002. *Analog Days: The Invention and Impact of the Moog Synthesizer.* Cambridge, MA: Harvard University Press.

Piore, M.J. and Sabel, C.F. 1984. *The Second Industrial Divide.* New York: Basic Books.

Polanyi, M. 1962. *Personal Knowledge.* London: Routledge

Polanyi, M. 1967. *The Tacit Dimension.* New York: Anchor Books.

Popper, K. 1959. *The Logic of Scientific Discovery.* London: Hutchinson.

Porter, T.M. 1995. *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life.* Princeton, NJ: Princeton University Press.

Price-De Solla, D.J. 1965. 'Networks of Science Papers'. *Science.* 149: pp510-515.

Price-De Solla, D.J. 1984. 'The Science/Technology Relationship, the Craft of Experimental Science, and Policy for the Improvement of High Technology Invention'. *Research Policy.* 13: pp3-20.

Price-De Solla, D.J. and Beaver, D. 1966. 'Collaboration in an Invisible College'. *American Psychologist.* 21: pp1011-1018.

Rabinow, P. 1996. *Making PCR: A Story of Biotechnology.* Chicago, IL: Chicago University Press.

Radiohead. 1997. 'Karma Police'. *OK Computer.* Parlophone.

Rao, G. 2004. 'Bioinformatics – New Horizons, New Hopes'. *Drug Discovery and Development.* July Anniversary Supplement. pp18-21.

Ravetz, J.R. 1971. *Scientific Knowledge and its Social Problems.* Oxford: Clarenden Press.

Ravichandran, V. and Sriram, R.D. 2005. 'Towards Data Standards for Proteomics'. *Nature Biotechnology.* 23(3): pp373-376.

Reiner, R. 2000. *The Politics of the Police.* 3rd ed. Oxford: Oxford University Press.

Research Council United Kingdom (RCUK). 2008. *E-science.* Available at: <URL: http://www.rcuk.ac.uk/escience/default.htm> [Accessed 3 September 2008].

Rincon, P. 2006. 'Science Takes Stock After Clone Row'. *British Broadcasting Corporation.* 10 January 2006.
Available at: <URL: http://news.bbc.co.uk/1/hi/sci/tech/4578286.stm> [Accessed 19 December 2006].

Roberts, L.G. and Wessler, B.D. 1970. 'Computer Network Development to Achieve Resource Sharing'. In *Proceedings of AFIPS Spring Joint Computer Conference.* AFIPS Press. pp543-549.

Rosen, R. 1978. *Fundamentals of Measurements and Representation of Natural Systems.* North-Holland: Elsevier.

Rothman, R. 1972. A Dissenting View on the Scientific Ethos. *The British Journal of Sociology.* 23 (1): pp102-108.

Rubin, H.J. and Rubin, I.S. 1995. *Qualitative Interviewing.* Thousand Oaks, CA: Sage.

Russell, B. 2004. *History of Western Philosophy.* London: Routledge

Ryle, A.P., Sanger, F., Smith, L.F. and Kitai, R. 1955. 'The Disulphide Bonds of Insulin'. *Biochemical Journal.* 60: pp541-556.

Scheele, A.G. 1975. 'Two-Dimensional Gel Analysis of Soluble Proteins. Characterisation of Guinea Pig Exocrine Pancreatic Proteins'. *Journal of Biological Chemistry.* 250: pp5375-5385.

Schiebinger, L. 1996. 'The Loves of the Plants'. *Scientific American.* Vol 274 (2): pp110.

Schiebinger, L. 2000. 'Taxonomy for Human Beings'. In Kirkup, G. Janes, L. Woodward, K. and Hovenden, F. eds. *The Gendered Cyborg: A Reader.* London: Routledge. pp11-37.

Schutz, A. 1943. 'The Problem of Rationality in the Social World'. *Economia.* 10(38): pp130-149.

Scott, S. 2007. 'Theatres of Talk: The Interviewer as Stagehand'. *Qualitative Researcher.* 5: pp6-8.

Sennet, R. 2008. *The Craftsman.* London: Allen Lane.

Service, R.F. 2001. 'High-speed Biologists Search for Gold in Proteins: Proteomics Aims to Chart the Ebb and Flow of Tens and Thousands of Proteins at Once to Produce Snapshots of Life Inside Cells. For Now, the Technology Isn't There. But this Young Field is Growing up Fast.' *Science.* 294 (5549): pp2074-2079.

Shapin, S. 1988. 'The House of Experiment in Seventeenth-Century England'. *Isis.* 79: pp373-404.

Shapin, S. 1996. *The Scientific Revolution.* Chicago, IL: University of Chicago Press.

Shapin, S. and Schaffer, S. 1985. *Leviathan and the Air-Pump: Hobbes, Boyle, and the Experimental Life.* Princeton, NJ: Princeton University Press.

Shapiro, R. 1992. *The Human Blueprint.* New York: Bantam.

Shrager, J. 2008. 'The Evolution of BioBike: Community Adaptation of a Biocomputing Platform'. *Studies in History and Philosophy of Science.* 38(4): pp642-656.

Smith, L.M., Sanders, J.Z., Kaiser, R.J., Hughes, P., Dodd, C., Connell, C.R., Heiner, C., Kent, S.B.H. and Hood, L.E. 1986. 'Fluorescence Detection in Automated DNA Sequence Analysis.' *Nature.* 321: pp674-679.

Smith, T.F. 1990. 'The History of Genetic Sequence Databases'. *Genomics.* (6): pp701-707.

Spradley, J. 1979. *The Ethnographic Interview.* New York: Holt, Rinehart and Winston.

Star-Leigh, S. 1991. 'Power, Technologies and the Phenomenology of Conventions: On Being Allergic to Onions'. In Law, J. ed. *A Sociology of Monsters: Essays on Power, Technology and Domination.* London: Routledge. pp26-56.

Star-Leigh, S. and Griesemer, J.R. 1989. 'Institutional Ecology, Translations and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology 1907-39'. *Social Studies of Science.* 19: pp387-420.

Stent, G.S. 1971. *Molecular Genetics: An Introductory Narrative.* San Francisco, CA: W.H. Freeman and Company.

Stephens, N. 2005. *Why Macroeconomics Orthodoxy Changes so Quickly: The Sociology of Scientific Knowledge and the Phillips Curve.* Ph.D. Thesis: Cardiff University.

Stephens, N. 2007. 'Collecting Data from Elites and Ultra Elites: Telephone and Face-to-Face Interviews with Macroeconomists'. *Qualitative Research.* 7(2): pp203-216.

Stephens, N. Atkinson, P.A. and Glasner P. 2006. *Bridging Strategies and Husbandry in the UK Stem Cell Bank* paper presented at the European Association for the Study of Science and Technology, 26 August 2006. Lausanne.

Stephens, N. Atkinson, P.A. and Glasner P. 2007. *Embodiment, Space and Representation in the UK Stem Cell Bank* paper presented at the British Sociological Association Conference, 13 April 2007. London.

Stephens, N. Atkinson, P.A. and Glasner P. 2008a. 'The UK Stem Cell Bank: Securing the Past, Validating the Present, Protecting the Future'. *Science as Culture.* 17: pp43-56.

Stephens, N. Atkinson, P. and Glasner, P. 2008b. 'The UK Stem Cell Bank as Performative Architecture'. *New Genetics and Society*, 27 p87-99.

Stephens, N. and Lewis, J. 2008. *Closing the Regulators Regress: Making Things Work in UK Stem Cell Science* paper presented at the European Association for the Study of Science and Technology, 23 August 2008. Rotterdam.

Stivers, R. 2001. *Technology as Magic: The Trimuph of the Irrational.* New York: Continuum Publishing Company.

Strachan, T. and Read, A.P. 1999. *Human Molecular Genetics 2.* Oxford: BIOS Scientific Publishers Ltd.

Strauss, A.L. 1987. *Qualitative Analysis for Social Scientists.* Cambridge: Cambridge University Press:

Suchman, L.A. 1987. *Plans and Situated Actions: The Problem of Human-Machine Communication.* Cambridge: Cambridge University Press.

Tanaka, K and Busch, L. 2003. 'Standardization as a Means for Globalizing a Commodity: The Case of Rapeseed in China'. *Rural Sociology.* 68 (1): pp25-45.

Taylor, C.F. 2008. 'Promoting Coherent Minimum Reporting Guidelines for Biological and Biomedical Investigations: The MIBBI Project'. *Nature Biotechnology.* 26(8): pp889-896.

Taylor, C.F., Hermjakob, H., Julian, R.K., Garavelli, J.S., Aebersold, R. and Apweiler. R. 2006. 'The Work of the Human Proteome Organisation's Proteomics Standards Initiative (HUPO PSI)' *OMICS: A Journal of Integrative Biology.* 10(2): pp145-151.

Taylor, C.F., Paton, N.W., Garwood, K.L., Kirby, P.D., Stead, D.A., Yin, Z., Deutsch, E.W., Selway, L., Walker, J., Riba-Garcia, I., Mohammed, S., Deery, M.J., Howard, J.A., Dunkley, T., Aebersold, R., Kell, D.B., Lilley, K.S., Roepstorff, P., Yates, J.R., Brass, A., Brown, A.J., Cash, P., Gaskell, S.J., Hubbard, S.J. and Oliver, S.G. 2003. 'A Systematic Approach to Modelling, Capturing and Disseminating Proteomics Experimental Data'. *Nature Biotechnology.* 21: pp247-254.

Taylor C.F., Paton, N.W., Lilley, K.S., Binz, P.A., Julian, R.K. Jr., Jones, A.R., Zhu, W., Apweiler, R., Aebersold, R., Deutsch, E.W., Dunn, M.J., Heck, A.J., Leitner, A., Macht, M., Mann, M., Marten,s L., Neubert, T.A., Patterson, S.D., Ping, P., Seymour, S.L., Souda, P., Tsugita, A., Vandekerckhove, J., Vondriska, T.M., Whitelegge, J.P., Wilkins, M.R., Xenarios, I., Yates, J.R., Hermjakob, H. 2007. 'The Minimum Information About A Proteomics Experiment (MIAPE)'. *Nature Biotechnology.* 25(8): pp887-93.

Taylor, F.W. 1911. *The Principles of Scientific Management.* New York: Harper Bros.

Thomson, J.J. 1899. 'On the Mass of the Ions in Gases at Low Pressures'. *Philosophical Magazine.* 48(291): pp547-567.

Thrift, N. 2006. 'Re-inventing Invention: New Tendencies in Capitalist Commodification'. *Economy and Society.* 35 (2): pp279-306.

Timmermans, S. and Berg, M. 1997. 'Standardization in Action: Achieving Local Universality through Medical Protocols'. *Social Studies of Science.* 272: pp273-305.

Timmermans, S. and Berg, M. 2003. *The Gold Standard: The Challenge of Evidence-based Medicine and Standardization in Health Care.* Philadelphia: Temple University Press.

Tiselius, A. 1937. 'A New Apparatus for Electrophoretic Analysis of Colloidal Mixtures'. *Transactions of the Faraday Society.* 33: pp524-531.

Toulmin, S. 1970. 'Does the Distinction Between Normal and Revolutionary Science Hold Water?' In Lakatos, I. and Musgrave, A. eds. *Criticisms and the Growth of Knowledge.* Cambridge: Cambridge University Press. pp39-47.

Tournay, V. 2008. 'From Standards to Concerted Programs of Collective Action: The Standardization Process of Cell Therapy in France. Pragmatist Contribution to the Sociology of Innovation'. *Social Theory and Health.* 6: pp220-238.

Travis, A. 2006. 'Clarke to Overhaul Drug Classifications'. *Guardian Newspaper.* Available at: <URL: http://www.guardian.co.uk/politics/2006/jan/20/drugsandalcohol.immigrationpolicy> [Accessed 11 February 2007].

Traweek, S. 1988. *Beamtimes and Lifetimes: The World of High Energy Physicists.* Cambridge, MA: Harvard University Press.

Turner, V. 1967. *The Forest of Symbols.* Ithaca: Cornell University Press.

Twain, M. 1883. *Life on the Mississippi.* Montreal: Dawson Brothers.

Twyman, R.M. 2004. *Principles of Proteomics.* Abingdon: Garland Science/BIOS Scientific Publishers.

Tyers, M. and Mann, M. 2003. 'From Genomics to Proteomics'. *Nature.* 422: pp193-197.

University of Bristol. 2007. *New Matrix of Harm for Drugs of Abuse'.* Available at: <URL: http://www.bristol.ac.uk/news/2007/5367.html> [Accessed 18 September 2007].

Utterback, J.M. 1996. *Mastering the Dynamics of Innovation.* Boston, MA: Harvard Business School Press.

Vallance, R. 2001. 'Gaining Access: Introducing Referred Approval'. *Issues in Educational Research.* 11(2): pp65-73.

Van Maanen, J. 1987. *Qualitative Methodology.* London: Sage.

Van Den Daele, W. and Weingart, P. 1976. 'Resistance and Receptivity of Science to External Direction: Emergence of New Disciplines Under the Impact of Science Policy'. In Lemaine, G., Macleod, R., Mulkay, M. and Weingart, P. eds. *Perspectives on the Emergence of Scientific Disciplines.* Chicago: Aldine Publishing Company. pp247-275.

Venter, C. *et al.* 2001. 'The Sequence of the Human Genome'. *Science.* 291: pp1304-1351.

Wade, N. 1981. 'The Complete Index to Man'. *Science.* 211 (4477): pp33-35.

Wainwright, S., Williams, C., Michael, M., Farsides, B. and Cribb, A. 2007. 'Remaking the Body? Scientists' Genetic Discourses and Practices as Examples of Changing Expectations on Embryonic Stem Cell Therapy for Diabetes.' *New Genetics and Society*, 26(3): pp251-268.

Waller, J. 2002. *Fabulous Science: Fact and Fiction in the History of Scientific Discovery.* Oxford: Oxford University Press.

Watson, J.D. 1968. *The Double Helix.* London: Wiedenfeld and Nicolson.

Watson, J.D. 1990. 'The Human Genome Project: Past, Present and Future'. *Science.* 248(495): pp44-51.

Watson, J.D. and Crick, F.H.C. 1953. 'A Structure for DNA'. *Nature.* 171: pp737-738.

Webb, E.J., Campbell, D.T., Schwartz, R.D., Sechrest, L. and Grove, J.B. 1966. *Nonreactive Measures in the Social Sciences.* Chicago, IL: Rand McNally.

Weber, M. 1968. *Economy and Society – An Outline of Interpretive Sociology.* New York: Bedminster Press.

Weber, R. 1990. *Basic Content Analysis.* London: Sage.

Webster, A. 2007a. *Social Science and Stem Cells Research: Key Aspects and Implications of the ESRC's Programme.* BSA Annual Conference, 12 April 2007. London.

Webster, A. 2007b. 'Reflections on Reflexive Engagement: Response to Nowotny and Wynne'. *Science, Technology and Human Values.* 32(5): pp608-615.

Weinberg, A.M. 1967. *Reflections on Big Science.* Oxford: Pergamon Press.

Weiners, B. 2002. *Making Headlines in 10,000-Point Type.* Available at: <URL http://www.wired.com/wired/archive/10.12/headlines.html> [Accessed 15 May 2006].

Welsh, E., Jirotka, M., Gavagham, D. 2006. 'Post-genomic Science: Cross-disciplinary and Large-scale Collaborative Research and its Organizational and Technological Challenges for the Scientific Research Process'. *Philosophical Transactions of the Royal Society.* 364: pp1533-1549.

Wilkins, M.R., Sanchez, J.C., Gooley, A.A., Appel, R.D., Humphrey-Smith, I., Hochstrasser, D.F. and Williams, K.L. 1996a. 'Progress with Proteome Projects: Why all Proteins Expressed by a Genome Should be Identified and How to do it'. *Biotechnological Genetic Engineering Review.* 13: pp19-50.

Wilkins, M.R., Sanchez, J.C., Williams, K.L. and Hochstrasser, D.F. 1996b. 'Current Challenges and Future Applications for Protein Maps and Post-Translational Vector Maps in Proteome Projects'. *Electrophoresis.* 17(5): pp830-838.

Williams, G. 1984. 'The Genesis of Chronic Illness: Narrative Reconstruction'. *Sociology of Health and Illness.* 6(2): pp175-200.

Wolpe, P.R. and McGee, G. 2001. 'Expert Bioethics as Professional Discourse: The Case Study of Stem Cells. In Holland, S., Lebacqz, K. and Zoloth, L. eds. *The Human Embryonic Stem Cell Debate: Science, Ethics and Public Policy.* Cambridge, MA: MIT Press. pp185-196.

Xiong, J. 2006. *Essential Bioinformatics.* Cambridge: Cambridge University Press.

Zadeh, L.A. 1965. 'Fuzzy Sets'. *Information and Control.* 8: pp338-353.

Ziman, J. 1984. *An Introduction to Science Studies: The Philosophical and Social Aspects of Science and Technology.* Cambridge: Cambridge University Press.

APPENDIX ONE

LIST OF INTERVIEWEES

NAME	AFFILIATION	GENDER	RELEVANCE
Dr. Harrison	Research Council	Male	Deputy Director for Science and Technology
Dr. Francis	EBI	Male	Developer of Database for Protein Modifications.
Dr. Phillips	EBI	Female	Senior Scientific Database Curator
Dr. Simmonds	EBI	Male	Senior Software Engineer
Dr, Green	EBI	Male	Head of Proteomics Service
Mr. Bond	EBI	Male	Industry Programme Coordinator
Mrs. Eaton	EBI	Female	Member of Outreach Team
Ms. Porter	EBI	Female	Member of Outreach Team
Dr. Campbell	University A	Male	Academic Lead in Proteomics Facility
Dr. Illingworth	University A	Female	Reader in Neuropsychiatric Genetics with particular interest in Bioinformatics.
Dr. Edwards	University A	Male	Professor in Molecular Cell Biology with particular interest in Proteomics.
Dr. Fairbrother	University A	Male	Research Fellow in Bioinformatics
Dr. Nielson	University A	Male	Professor and Member of the European Standardisation Committee for the Implementation and Analysis of Gene Expression data.
Dr. Johnson	University A	Male	Doctor in Bioscience with particular interests in Bioinformatics and Microarrays.
Dr. Kennedy	University A	Male	Lecturer in Biomedical Informatics.
Dr. Cherry	University A	Male	Senior Lecturer in Biochemistry. Interests in Bioinformatics and Proteomics.
Dr. Elias	University A	Male	Doctor in Medicine developing Lymphocyte Nuclear Proteomics
Dr. Harris	University A	Male	Lecturer in Protein Science
Dr. Daniels	University A	Female	Bioinformatician
Dr. Andrews	University A	Male	Experimental Biologist and Bioinformatician.
Dr. Kenwood	University A	Male	Professor of Pathology with particular interest in Bioinformatics
Mr. Jenkins	University A/B	Male	Researcher and PhD student in Bioinformatics

Mrs. Wiley	**University B**	**Female**	**Database Manager**
Dr. Davies	**University B**	**Male**	**Lecturer in Genetics. Interests in Proteomics and Transcriptomics**
Dr. Bunn	**University B**	**Male**	**Senior Lecturer in Medical Biochemistry Genome and Structural Bioinformatics**
Dr. Strauss	**University C**	**Female**	**Director of Proteomics Facility**
Dr. Morris	**University C**	**Male**	**Professor of Statistics with particular interest in Proteomics**
Dr. Matthews	**University D**	**Female**	**Director of Proteomics Facility**
Dr. Griffiths	**University E**	**Male**	**Reader in Faculty of Life Sciences working in Bioinformatics and Proteomics**
Dr. Hardwick	**Not Affiliated**	**Female**	**Ex EBI Scientific Database Curator**

Main E-mail Contributions

Dr. Llewellyn	**University E**	**Male**	**Professor of Chemistry and Head of Interdisciplinary Centre**
Dr. Harvey	**University F**	**Male**	**Director of Systems Biology**
Dr. Evans	**University G**	**Male**	**Managing Director of Cell Biology Research**

APPENDIX TWO

LIST OF ACRONYMS AND ABBREVIATIONS

2DE Two-Dimensional Electrophoresis
2DPAGE Two-Dimensional Polyacrylamide Gel Electrophoresis
A Alanine
ANT Actor Network Theory
ARPA Advanced Research Project Agency
BBN Bolt, Beranek and Newman
BBSRC Biotechnology and Biological Sciences Research Council
BBU Bioinformatics and Biostatistics Unit
BLAST Basic Local Alignment Search Tool
BS British Standards
BSA British Sociological Association
C Cysteine
CERN European Centre for Nuclear Research
CESAGEN Centre for the Economic and Social Aspects of Genomics
CIHR Canadian Institute of Health Research
DARPA Defence Advanced Research Project Agency
DDBJ DNA Databank of Japan
DOE Department of Energy
DNA Deoxyribonucleic Acid
DSK Dvorak Simplified Keyboard
DTI Department of Trade and Industry
EBI European Bioinformatics Institute
EGENIS Centre for Genomics in Society
ELSI Ethical, Legal and Social Issues
EMBL European Molecular Biology Laboratory
ENMBnet European Biology Network
ENIAC Electrical Numerical Integrator And Calculator
ESI/MS Electrospray Ionization Tandem Mass Spectrometry
ESCR Economic and Social Research Council
ExPASy Expert Protein Analysis System
FFE Free Flow Electrophoresis
G Glycine
GENBANK Genetic Sequence Data Bank
GC/MS Gas Chromatography Mass Spectrometer
GCSE General Certificate of Secondary Education
GPS General Proteomics Standard
HEP High Energy Physicists
HGP Human Genome Project
HPI Human Protein Index

HPP Human Proteome Project
HPS History and Philosophy of Science
HTML HyperText Mark-up Language
HUGO Human Genome Organisation
HUPO Human Proteome Organisation
HUPOST Human Proteome Organisation newsletter
INNOGEN Centre for the Social and Economic Research on Innovation in Genomics
ISB Institute of Systems Biology
ISO International Organisation for Standardisation
MALDI-TOF/MS Matrix Assisted Laser Desorption/Ionisation Time-Of-Flight
MCP Molecular Cellular Proteomics
MIAME Minimum Information About a Microarray Experiment
MIAPE Minimum Information About a Proteomics Experiment
MGED Microarray Gene Expression Database
MOLGEN Molecular Genetics
MRC Medical Research Council
mRNA Messenger Ribonucleic Acid
MS Mass Spectrometer
NAS National Academy of Sciences
NCBI National Centre for Biotechnology Information
NIH National Institute of Health
NIHBDC National Institute of Health Bioinformatics Definition Committee
NHRGI National Human Genome Research Institute
NLM National Library of Medicine
NSF National Science Foundation
OLS Ontology Lookup Service
PCR Polymerase Chain Reaction
PIR Protein Information Resource
PP Interaction Protein-Protein Interaction
PRIDE Protein Identification Database
PSI Proteomics Standards Initiative
PSI-MI Proteomics Standards Initiative – Molecular Interactions.
QI Qualitative Inquiry
QTOF/MS Quadruple Time of Flight Mass Spectrometer
QWERTY Name derived from first six characters in the far left of the top row of keyboard
RAND Research and Development
RESID Protein Modifications Database hosted by EBI.
RNA Ribonucleic Acid
RSSDP Research Students' Skills Development Programme
RTF Rich Text File

PROSITE Database of protein domains, families and functional sites
RCUK Research Councils United Kingdom
SB Systems Biology
SIB Swiss Institute of Bioinformatics
SEE Studies of Expertise and Experience
SEQ Safety Efficacy Quality
SoCo Southern Comfort
SSK Sociology of Scientific Knowledge
SSS Social Studies of Science
STS Science and Technology Studies
T Threonine
TEA Laser Transversely Excited Atmospheric
TIGR The Institute for Genome Research
TrEMBL Translated EMBL
UK United Kingdom
UN United Nations
UNSW University of New South Wales
US United States
USSR Union of Soviet Socialist Republics
WWII World War Two
WWW World Wide Web