The Archive Saga

Shepherds of data, documents and code, and their will to order

Kristrún Gunnarsdóttir

Thesis examiners:
Professor Graham Button (external)
Professor Tom Horlick-Jones (internal**)**

Cardiff University

Submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

Minor amendments were made for this online dissemination, Oct. 2010.

...þetta verk er tileinkað ömmu og afa

Acknowledgements

This thesis came to fruition with the generous support of many: my colleagues and faculty at Cardiff and Cornell Universities and the National Library of Iceland - my family and friends.

First I like to pay my respects to those who brought me into the world of digital library developments, Þorsteinn Hallgrímson and Thomas Hickerson. They gave me challenging tasks, some of which I certainly struggled with. But their belief in my abilities to solve problems, and *put things to work*, opened the gateways to the grounds on which I stand today. I am indebted to my colleagues and co-workers at the two libraries, Cornell and the National Library of Iceland, with whom I settled many differences in our otherwise productive interdisciplinary collaborations. I pay my respects to Paul Ginsparg who agreed on a development of a doctoral thesis about the archive while I was still at Cornell, and with whom I had many thoughtful exchanges about the apparatus and its potential futures. My co-workers on the archive also deserve the sincerest acknowledgements for their patience under trying conditions in an environment filled with both amusements and bemusements.

I have come a long way developing this thesis but with the generous support of faculty in the School of Social Sciences, I now see this project to an end. I thank first Robert Evans and Ian Welsh who both have provided invaluable support and I would not have done this without them. I thank Neil Selwyn for his sharp mind and directness, Harry Collins for being himself, and the heads of department for financial support. I would also like to thank Trevor Pinch, in the Department of ST&S at Cornell, who supported this thesis in the early days, and Michael Lynch, who unwittingly won me over to ethnomethodology.

I greatly appreciate the assistance provided by my parents, Gunnar and Guðfinna, and the sponsorship granted by my uncle, Þorleifur. Thorsten Schwander deserves special thanks for his many lessons in hacking and for the times we spent hiking and chatting. My siblings, Gunnar and Helga, have been entertaining hosts and good company (along with spouses) on my recreational trips during this time, and my friends in Cardiff, Bruce, Mike, Ingrid, Merryn, Stuart, and many great others, provided enlightening discussions and good times. I also reserve gratitude to my new colleagues at Lancaster University who are exceptionally creative and intellectually stimulating, and gave a second wind to my efforts to finish this thesis. Finally, I reserve the deepest gratitude to Miguel. He has `poked and prodded' at my thesis, helped me to overcome many of the barriers in its development and seen me through the darkest moments. He has been my lifeline, intellectual partner and best friend.

Abstract

This thesis is an enquiry into expert-intensive work to support the aggregation and mass-dissemination of scientific articles. The enquiry draws on computer-supported actions and interactions between the people who worked as daily operators, and submitters who worked the system from remote as end-users. These recorded actions and interactions render visible matters of practical responsibility and competence with reference to reportable-accountable use of objects and devices. As a scholarly contribution, this thesis draws on the work started by Garfinkel and his followers, ie., the study programme of ethnomethodology. It poses a simple question—how do the people involved in these computer-supported actions and interactions do just what they do? In the attempt to answer that question, the enquiry shows that a competent and accountable use of system supports is manifested in reference to particular phenomena that need constant management—eg. warnings, anomalies, unknown entities, failing processes, disorderly work objects, and spurious actions—in short, phenomena of *disorder*. Records of how these phenomena are detected and attended to, and how particular problems are solved, reveals a complex relationship between computational functions and subtle human judgements. Without making generalised theoretical claims about this relationship, I conclude that remarkably little is known about the actual lived work that takes place in the operation and use of computer systems that are worked specifically to detect and cope with `anomalies' and thus require detective and reactionary labour. I offer this enquiry into the goings-on at this particular site as a singular opportunity to respecify *problems of disorder*.

Table of Contents

Chapter 1

Introduction

On the face of it

This thesis is about the everyday management of complex virtual objects, namely, electronic documents and records, and the administrative environment to process them. I shall simply called it `the archive'. On the face of it, this was an online preprint service with two basic features: the uploading and downloading of scientific articles. The clientèle were, for the most part, physicists and mathematicians looking for new preprints in their field or submitting their own. On the back end however, operators provided the necessary labour to support this throughput and ensure a robust and consistent service.

At the time when I first came on board, in October 2001, I was made aware of the amount of articles archived every day. It averaged close to 140 new submissions during my time on board,[1] with only one or two operators supporting the throughput on the back end. Monthly downloads were measured in hundreds of thousands.[2] In short, this system was a major storehouse of scientific works under strain by frequent hits and visits.

For the end-user, an important feature of this service is its accessibility on the internet. The archive has been considered a powerful exchange tool for scientific research for two main reasons. Firstly, there is significant time lapse between preprints and journal prints and, as advocates have argued, the quality of preprints is sufficient for ongoing research (Ginsparg, 1996; Kuperberg, 2002). Secondly, access to t he articles is free of charge. This means global access without `tollgates' (Harnad, 2001; Bachrach et al, 1998).

The fact that the articles are preprints and not peer-reviewed publications follows a long

1 New submissions averaged per day at roughly 139 in the year 2002 (from automated logging).
2 This figure also applies to the year 2002, based on figures provided by the inventor of the system.

tradition of preprint exchange in the particle physics community (O'Connell, 2002). There are well known debates about openly circulating this literature, there are questions about the focus and role of peer-review in physics and mathematics, and what impact the preprint exchange has had on scholarly communication (Odlyzko, 2000; Odlyzko, 2003). I have dealt with this matter elsewhere (Gunnarsdóttir, 2005). My central concern here is the everyday mass-dissemination of electronic documents and reports, and my interest in this aspect of the operation can be briefly explained as follows. I learned that the operators, and a few of their colleagues, had been working to develop the facility - and working there as operators themselves - for just about a decade. I want to show the work that took place in the operation and use of the archive and consider my enquiry in relation to ethnomethodological studies of action, order and reasoning, and the use of computer supports.

The back end

On my first visit to an operator's office, the operator drew sketches on large sheets of white paper to illustrate the flow of data packets, time constraints, points of inspection and interception. Then my colleague and I watched him work for a while at a terminal, negotiating with bits and pieces of software about the integrity of data components: formats, processing successes and failures, and so on.

My first impression was how hard it was to grasp what was going on. I was asked to express my thoughts on this operation to senior library personnel and I drew a sketch to broadly depict the activities based on my initial impression (Figure 1.1). What this sketch was meant to illustrate at the time was the locus of administrative control and how the contributions of everyone involved, together with the computer supports, fitted into that picture. The readers (**R**) could access the articles by following links provided on the archive website or by using the online search facility. Submitters (**S**) were provided with tools to upload their preprint source files and they received, by email, all sorts of automated

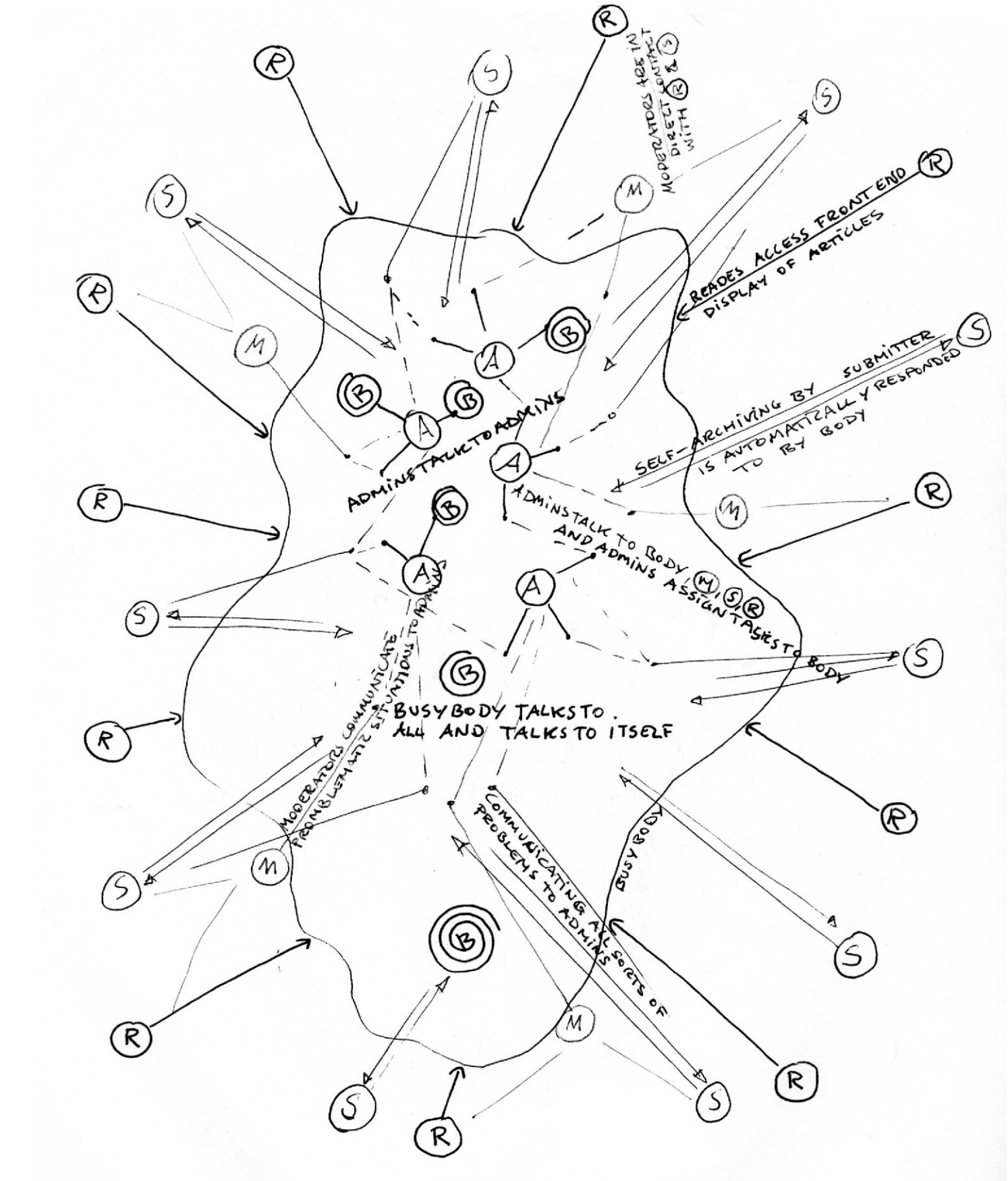

Figure 1.1: *My sketch of activities from October 2001. I distinguished readers of scientific documents* (**R**) *from submitters* (**S**) *to clarify that downloading a preprint and uploading one are two very different activities.* **M** *represents moderators who take a quick glance at incoming preprints.* **A** *represents administrators and* **B** *the computational devices that comprise `busybody'. The sketch indicates that* **M** *are in direct contact with* **S** *or* **R**. *This was misunderstanding on my part.*

responses which they sometimes had to act on to solve a problem with the processing of their submissions. Moderators (**M**) would access new preprints before they were made publicly accessible. They made suggestions if an article was not, as they put it, `topic relevant'. Admins (**A**) were the key operators, at the heart of the operation, so to speak. Although new submissions were shipped through with auto-processing supports - referred to as `busybody' (**B**) - operators were using hands-on controls to monitor the throughput and the work of busybody at various stages. The communication, controls and commands, and the preprint objects themselves, were predominantly in textual format. So, orienting and working in this environment, as I was to find out, demanded advanced computer literacy. This applied not only to the operators but also to submitters. I considered myself a literate so it turned out that as a member of a library systems team I also became an operator on the archive.

A depiction of housekeeping

The archive is an academic resource service with intimate ties to librarianship and the development of digital libraries. The facilities are an amalgam of computerised document / communication / operations technologies assembled to aggregate and disseminate scientific literature. The operation and use of the archive has been associated with several interrelated histories which include those of: (1) document and records management; (2) data transmission and database engineering; (3) scholarly writing and research libraries, and (4) scientists' methods of representing formulas, symbols and charts. The key historical landmarks, typically mentioned, are the advent of mass printing and the first scientific journals (Knorr-Cetina, 1981; Harnad, 1991; Guédon, 2001).

The aggregation and dissemination of research results became an important feature in the mediation of scientific research and reasoning in the 17th and 18th centuries (Schaffer, 1998). Newly founded science societies established committees on literature and publications. Dissemination through mass printing was a means to further broaden the circle of expert witnessing to scientific experimentation, and catalogues of records and

retrospective reviews are early examples of so-called aggregation techniques. To the present day, individual scientists, science societies, research institutions and government bodies, have been active as publishers, collectors and managers of specialised archives, issuing catalogues and reference materials (eg. Adkinson, 1976; Werdel and Adams, 1976; Miles, 1992; historical overview by Guédon, 2001). This trend continued after computers became tools to process information (Mooers, 1950). For example, physicists and mathematicians began undertaking pilot projects using records from collections of their own works (discussion in Swanson, 1988; also overview of information retrieval experiments in Jones, 1981). But the development of computer technologies to do what was already accomplished with the existing techniques took decades to deliver instruments and services for everyday use in the scientific communities.

Physicists and mathematicians saw gradual changes. The Linear Acceleration Center at Stanford University (SLAC) and the Deutsches Elektronen Synchrotron (DESY) in Hamburg were collecting all journal articles and preprints circulating among particle physicists and associated colleagues as early as 1962 (O'Connell, 2002). They maintained bibliographic records in card catalogue but DESY also hired physicists to assign keywords to classify the records. Then the SLAC catalogue became a test subject for the Stanford Physics (later Public) Information REtrieval System (SPIRES) in 1968, and the American Physical Society in collaboration with SPIRES started a weekly publication of records printed from this database, known as Preprints in Particles and Fields. The two libraries, SLAC and DESY, began collaborating on these publishing efforts and, by 1974, scientists on both sides of the Atlantic were subscribing to regular listings of new preprints and journal publications with keywords and citation overviews which were sent out to subscribers by ordinary mail service.

With the widespread use of the IBM mainframe for electronic file and message communications (see Abbate, 1999), `bitnet` became the established protocol to operate online subscriptions to this records database. Finally, a significant development occurred in the late 1980s when a scientific typesetting system, `TeX`, was made freely available by the

computer scientist Donald Knuth.[3] It quickly became popular among mathematicians and physicists who could now produce a scientific document on a Unix desktop of equal presentational quality to a published article. Prior to that, only print shops could print symbols, formulae and data charts for scientific publications.

Operating an archive

It is specific to accounts of the origins of the archive that bitnet, TeX typesetting technology and the tradition of preprint exchange are seen as central to its installation in 1991. The key concept was, as Ginsparg has put it, *a pure dissemination system* (Ginsparg, 2003). A group of less than 200 theorists in high energy physics were working on a matrix model of string theory and two dimensional gravity when one of them offered his Unix office computer for the exchange of new preprints typeset in TeX. In the first two years this office computer served as a simple email reflector for the uploads and downloads of TeX source files. The custom designed software also sent listings of new submissions to subscribers by email. In other words, a simple self-service was set up and a centralised repository of electronic preprints began to grow.

The use of the archive has spread consistently among TeX and Unix literate physicists and mathematicians (Figure 1.2).[4] Similar initiatives appeared in the 1990s to encourage scholars to archive online and to make the academic literature available free of charge (Okerson and O'Donnell, 1995). In principle, anyone could post documents at a minimum cost and build their own repositories. General purpose software was even introduced to facilitate such efforts and so-called `open archives' were set up.[5] But online self-archiving and digital library building invite complications. One problem often mentioned is that IT

[3] Donald Knuth first introduced TeX in an invited lecture at the American Mathematical Society (AMS) in 1978, but it took another decade to develop an industrial strength product. See the AMS website on the Josiah Willard Gibbs lecture series at http://www.ams.org/meetings/gibbs-lect.html. There were other scientific typesetting and document processing programmes being developed at the time, for instance Troff, see http://en.wikipedia.org/wiki/Troff. See also Donald Knuth's web page, http://www-cs-faculty.stanford.edu/~knuth/, and about TeX at http://www.tug.org/.

[4] See Appendix III for a list of subject archives in the system, ordered by date of creation.

[5] The *eprints* software, for instance, was developed to do this, see http://eprints.org.

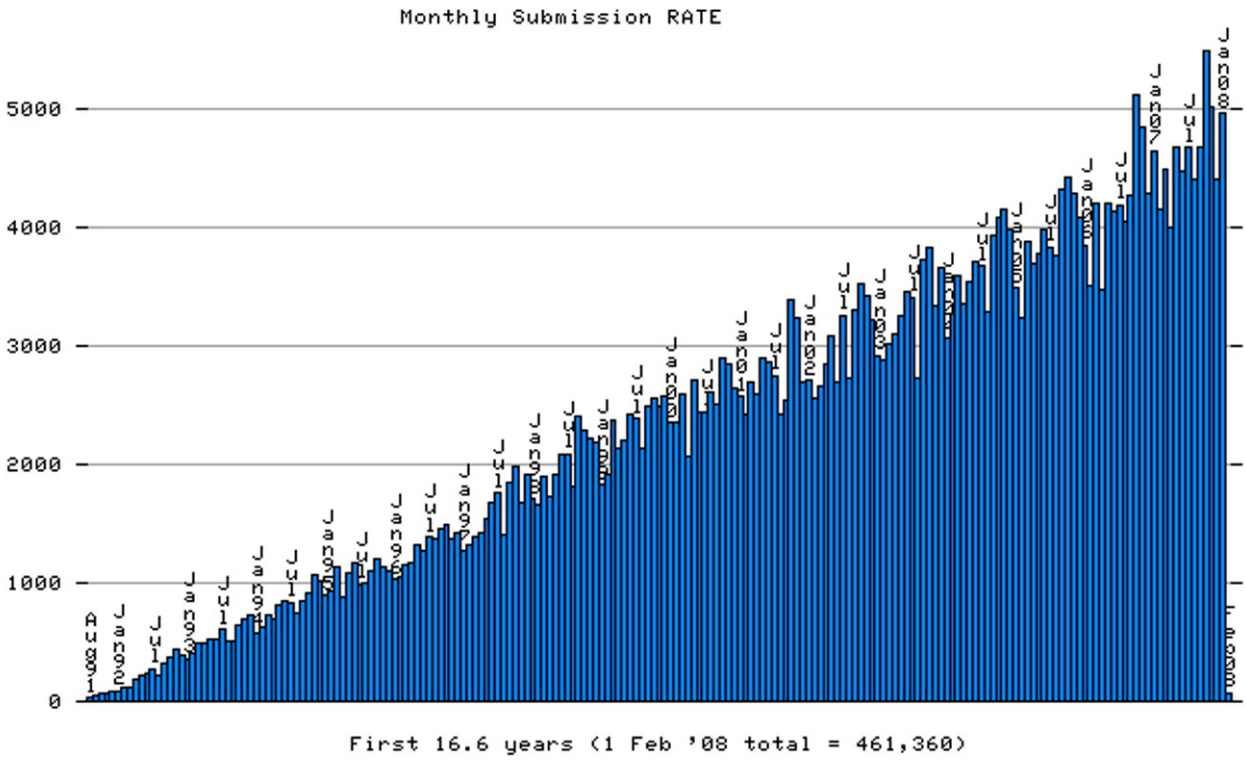

Figure 1.2: *Submission figures illustrated on the archive website*

literacy and computing / programming skills have an enormous range in and between departments of academic institutions. So, the *do-it-yourself* approach has generally not been attractive (Pinfield, 2001; Okerson and O'Donnell, 1995: ch.IX and X in particular on these complications). In this respect, the archive differs significantly from other forms of online publications (see also Bohlin, 2004).

There were no attempts to create `dialogues' between computers and novices in the design of interfacing features and other supports on the archive. The use of the TeX typesetting environment is a case in point. Almost all of the submitters were typesetting their articles in TeX but using the software requires literacy in the mark-up language and the engineering environment that renders the visual representations of articles into postscripts and PDF files [see TeX in Appendix I, glossary of terms]. Correspondingly, TeX know-how was required to use a TeX auto-compiler on the archive. For the operators however, one of

the main challenges was that conditions for using and operating the system would change over time in ways that was not so much in their control. They had to be sensitive to progressions and developments elsewhere—developments of document rendering and conversion tools, figure manipulation tools, and other software potentially useful for preparing, formatting, archiving and disseminating preprints. They considered it their job to stay abreast of such developments, incorporate new protocols and upgrade software. They were also active contributors to some of these progressions (Warner, 2001), but mainly their craftsmanship was put to use in accommodating and handling TeX, in handling figure manipulations, records processing and, more generally, coping with (and within) a virtual and distributed work environment. Doing this work in the operation and use of the archive is what I will enquire into, ie., these technologies *in action* as one could say (Heath and Luff, 2000).

Technology in action

I argue that studies in ethnomethodology provide an important resource for my investigation of how the people involved in the operation and use of the archive organise and do their work. My intention is to uncover spontaneously produced interactions by those who are all-in-one, technical administrators, system builders, daily operators and planners for the future. It is of equal interest to examine the interactions with the clientèle who work the system from remote as end-users. I will argue that the various ways in which the activities are structured have no place in normative reasoning about organisational conduct, nor in formal descriptions of agents, devices and goal-oriented tasks. I will also argue that the ways in which the system is worked has no place in social-historical reasoning about origins and circumstances of technical concerns and decision-making. The persons who work the system make visible their own critical and constitutive enquiries. These enquiries negotiate and explicate, locally and interactionally, both histories and futures, what the objects-at-hand are, and how this ongoing project, the archive, is worked in reference to issues of practical responsibility, legitimacy and competence, and how to keep things in order. But if studies in

ethnomethodology are an important resource for enquiring into the operation and use of technologies in this particular organisation, how is it that conventional social and historical studies or those based on cognitivist assumptions are not?

What the term, technology, stands for has been open to many interpretations. Sociologists have called it a slippery term (Pinch and Bijker, 1984), meaning, that the broad use of the term effectively obscures analytic distinctions between artefacts, concepts and practices. Such broad use has also served to make sweeping generalisations about technology as progressively prominent in our private, occupational and public lives. Progressions are cast as socially/culturally determined or determined by technological reasoning, or some mixture of the two. As Marc Berg has argued, opposing technological versus social determinism plays on dominant scholarly interests which are essentially about the ways in which humans are believed to be affected by technology: empowered or threatened, rewarded or disciplined (Berg, 1998, p.474). The presumed contrast between authoritarianism and democracy is also invoked to argue that if technologies are socially/culturally negotiable they will be better designed, for example, workplaces more democratic. As Berg puts it, such views concern the kinds of worlds we imagine technologies can support and how developments can be intercepted and affected to bring about more justly ordered collectives (1998, p.457).

It is common place in sociological studies that the logics of technical reasoning and technical work, remain *external* to reasoning and explanation for why a technology turns out as it does, why it has certain impacts and how it is impacted. Nonetheless, scientifically oriented technology and scientific methods provided the early moderns with the means to develop professional sociology—a systematic study of social or, say, social-scientific categories. What have been seen as the *internal* domains of technology were not considered proper sociological topics until the 1970s when a small group of social scientists and social historians began to include technical reasoning in a social-epistemological model of knowledge and know-how (eg. Pinch and Bijker, 1984). In other words, the logics internal to technical domains were now explained as thoroughly socially constructed.

What ethnomethodologists have pointed out with respect to these developments is how the phenomena of naturally unfolding situations, concerns and activities, are either trivialised or ignored by sociologists whether their explanations consider internal or external domains. As Michael Lynch has put it, investigators apply sociological and historical methods that reconcile actual situations, concerns and decisions, with the investigators' analytic definitions for social conditions and origins (eg. Lynch, 1982). Ethnomethodologists have also argued that formal descriptions such as task analysis, assumptions about cognitive control structures, and normative reasoning about the order of work, overlook the many ways in which work procedures and the use of technical systems are *ad hoc*, as Crabtree puts it, and are necessarily so in order to accomplish and coordinate work on an ongoing basis (Crabtree, 2001). In turn, ethnomethodologists have demonstrated how the enquirer can attend to naturally occurring actions and activities and, in doing that, enable a reproduction of what are intuitively recognisable practices *in-their-course*.

It is in light of these critiques and conventions that I will review some of the study policies and logics of sociological reasoning. I will also look at the ways in which technology and technical work has been theorised and I will examine some of the cognitivist assumptions that typically underpin the designs of computer systems and supports. I will show how a competent and accountable use of a particular system in an organisation is asserted and agreed upon by participants themselves in the sense that they see - more or less - where they are at and what to do next. These are substantive work results, achieved on an on-going basis, and inseparable from the local knowledge and reasoning involved in producing such results. So, this thesis will adopt an ethnomethodological dictum of enquiring into and describing the phenomena of the practical work achieved. It will follow a tradition of how ethnomethodologists have responded to cognitivism and conventional sociological attempts at solving the *problem of order*. My point of departure however, is that in doing so I will shift the attention to how the phenomena of *disorder* are worked in reference to issues of social relevance, and how to keep objects, devices and persons more or less on an even keel.

Why examine the phenomena of *disorder*?

Following Ulrich Beck's work on risk society (Beck, 1992), considerable attention has been drawn to conditions of uncertainty and risk, fragmentation and conflict (see also Giddens, 1999). Advanced technologies are certainly implicated in such problems as well as in managing uncertainties and security and maintaining *the order* (eg. Amoore, 2006; Allhoff et al, 2007; Aradau and Van Munster, 2007). We are also told that individuals *can* radically reshape and restructure their `multi-faceted' surroundings whereby self-assertive, innovative and technically-informed activities act on global forces and societal change. It may seem then that we should abandon the long standing agency-structure or actor-system dilemmas, and shift our focus to theorise complexity, the sources of disorder and the relationship between agency and perpetually uncertain societies. What I have in mind is rather different.

Jordan and Lynch (1992) describe conditions of instability and fragmentation in `routine' technical work as quite obvious to the technicians themselves as well as to those who use their products (Jordan and Lynch, 1992). What they also discover is that technical work can be mastered through a solitary practice and to `privatise' techniques leads to sporadic invasions of variants in what can be entirely glossed over as the *same* procedure or, in short, business as usual. But, with all the attention ethnomethodologists have given to the ways in which order is achieved and how business is as usual, remarkably little has been done in the way of treating sporadic invasions of variants, detection of anomalies or the fragmentation of method and procedures in *coping* as “pre-theoretical phenomena of considerable interest”, to use Jordan and Lynch's words (p.85).

In this thesis, I take the *problems of disorder* in technical work as my point of departure —a term that Jordan and Lynch refer to perhaps tongue-in-cheek. Computer technologies have been promoted and pushed through budget meetings and managerial offices with the argument that they will improve efficiency and productivity, even that they will eventually achieve the perfect idealised order via automation. For example, the system in question in

this thesis is described by its inventor as *fully automated* or, in any event, a system that *could* and *should* be fully automated. However, the evidence is that real-life settings—involving living, breathing persons going about their business—do not lend themselves to perfectly ordered and fully automated processing and control systems. Systems are usually support systems to human-controlled operations or so-called automated system actually need constant monitoring and maintenance to keep unforeseeable and disorderly entities and activities in check. For example, the people who took care of daily operations on the archive called it the daily grind. At least one person had to be on the look-out, so to speak, detecting, investigating and solving problems. One thing that could be taken for granted in doing this work was that end-users, working the system from remote, would do all sorts of unpredictable things when preparing and uploading their submissions. What could also be taken for granted was that data and document handling technologies are an open field of possibilities which present themselves sporadically, for example, when the system suddenly could not process an upload and finding the reason for that required investigation. The operators had to be vigilant and meticulously attentive to detail. It was necessary to take nothing for granted and be ready to investigate any anomalies as they emerged. Furthermore, there was great difficulty in specifying the job description of an operator.

We can argue then, with good reason, that such a system with which users and operators grapple, but normally cope, should be recognised as only relatively stable and only partially structured. It should also be recognised that the work that takes place is not orderly in any conventional sense which assumes that *normal routines* are achieved individually or collectively, or that tasks and skills are attributed without much ado. For example, when asked about the warning, `"*** PS BAD ***"` (a few every day), one operator explained to me that they were looking for *this, that* or *the other* possibility which could mean any number of things depending on such and such conditions. Then they could be dealt with in *this, that* or *some other* way depending on priority, preference and available techniques. In other words, most explanations I received when I first came on board began with remarks like “there are

no hard and fast rules" or "it depends". What I quickly learned was that working the daily grind was pervasively uncertain in the sense that contingencies, anomalies and exceptions, invading the scene, and invention of variants in coping with them, was what normally took place.

What I will do then is to foreground and enable the reader to see the phenomena of ongoing detective, reactionary and preventative labour in order to respecify, ethnomethodologically, the *problems of disorder* in the operation and use of the archive. But I should admit that speaking of order or disorder is somewhat confusing. The technical uses of these terms suggest that achieving theoretical soundness of *the order* (or otherwise *disorder*) will explain how coherent private, occupational or public activities are possible even if they are neither orderly nor disorderly in any theoretical or idealised sense. But non-technical uses of these terms also risk being misleading. Ethnomethodologists emphasise *order* as what is achieved and *normal* as business-as-usual. The terms are supposed to mean what they would mean to anyone. *Normal* is not a technical term to describe the outcome of normative reasoning. *Order* is not a technical term for an inductive problem of explaining action and the social order. But to plainly assume that there is *business-as-usual* in keeping things in check, detecting disorders and reacting to them, introduces a confusing set of issues. One can argue that there is nothing normal about how the archive worked and how it was worked by persons who laboured with inspection and detection, trial and error. Order is not exactly what is achieved either. Rather, what is visibly rational and functionally relevant to practical purposes is a *more or less* agreeable outcome on an open horizon of how *things* might play out in the future. The counter-argument is perhaps then to argue that if flux, variation, anomaly, conflict, and so on, are the objectives of everyday affairs, then that is normal. And, if the work of coping with such phenomena is business-as-usual then order is indeed what is achieved in keeping a system running relatively smoothly and more or less error-free. Are we then simply drawn into a scholarly project of describing what anyone can assume is normally recognised as orderly or disorderly? The *social order* is, after all, a

foundational matter of interest to conventional sociologists and ethnomethodologists alike, although they deal with the matter very differently as we shall see. What I propose is to use this enquiry to sort out more elaborate reflections on this matter.

The order of the thesis

The next three chapters will clarify my own relation to the `project of theorising' and explain some of the study interests of ethnomethodologists with reference to their critiques of sociological and cognitivist reasoning. Chapter 2 will address conventional sociology and review the contributions ethnomethodologists have made in their studies of social action, order and reasoning, technology and technical work. By comparing conventional studies on these topics with studies in ethnomethodology, I seek to clarify how it is that studies in ethnomethodology are an important resource for the development of this thesis. Chapter 3 will address common critiques of cognitivism in relation to studies that are specifically directed at the use of computer systems and computer supports in organisational interaction. Chapter 4 will consider social science research methods and clarify the study policies of ethnomethodology in reference to method and analysis developed in this thesis.

In Chapters 5 and 6, I describe and analyse how people involved in the operation and use of the archive do what they do. The analysis is divided into topics organised around emerging issues of *competence and accountability* in managing disorderly phenomena, namely, the administration of incomplete, incorrect or corrupted scientific articles. In doing this, I address separately some of the circumstances in which operators and submitters engage in practical responsibility, inspection, detection work and corrective action, as efforts to solve problems with submissions of preprints. Analysis of scenarios and events reveal a complex relationship between computational functions and subtle human judgements.

In Chapter 7, I will make the argument that the study interests of ethnomethodologists are relevant to professional practitioners whose actions and interactions are considered in this chapter. I will take examples of events in which operators, submitters and university library

staff were faced with challenges arising from incompatible tools and methods. From these conflicting expectation, I argue that much can be learned from attending to systems in real-life settings involving living breathing persons going about their business.

In Chapter 8, I will summarise my enquiries and draw concluding remarks about the implications raised in this thesis. I will argue that the manual support in the daily grind was the work of establishing deviations and anomalies and materialising a will to order. What we see was never supposed to take place—the everyday management of warnings, anomalies, unknown entities, failing processes, disorderly work objects or spurious actions. I will argue that the natural disorders of the world pose a particular challenge which is not given due attention and I will explain how it is that the study interests of ethnomethodologists concern matters we should want to know more about.

Chapter 2

Ethnomethodology and sociological reasoning

The Following sections will explain how it is that studies in ethnomethodology provide such an important resource for the development of this thesis. In doing this, I will first address the ways in which professional sociology is established as a scientific discipline and clarify what it is that ethnomethodologists criticise. I will review some of sociology's central challenges such as the agency versus structure debate and the distinction between internal and external explanations of technology and technical work. I will review how conceptions of technology have served to substantiate the study methods of sociology and, later, how the *epistemological contents* of technology became the topic of sociological reasoning. The chapter will conclude by clarifying my own relation to the project of reasoning and enquiry.

Ethnomethodology's critique

Garfinkel's campaign on the margins of sociology is conducted largely in response to sociology's disciplinary continuity as a systematic study of social-scientific categories. In his writings about studies in ethnomethodology, he criticises, among other things, the claims that were made on the basis of structural functionalism, notably Parsons' thesis (Garfinkel, 1988; Garfinkel, 1991; see also Garfinkel, 1967 and discussions in Button, 1991a; Lynch, 1993; Button and Sharrock, 1998; Dourish and Button, 1998). Structural functionalism however, is only one historical example of a long tradition of constructive analyses which draw on inspirations from the early modern theorists. What they all have in common, as Garfinkel points out, are the uses of systematically structured metalanguage which is applied to explain social action and the social order (Garfinkel, 1988). If such abstractions are taken seriously,

as objective theoretical principles in the formal analysis of order (action, work, etc.), Garfinkel argues that "[o]rderlinesses in the plenum pose for formal sociological analysis its tasks of detecting and specifying that orderliness and demonstrating it in massively recurrent, distinctive, essential, invariant identifying details of formally analyzed structures of practical action." (Garfinkel, 1988, p.106; see also Garfinkel, 1986).

Garfinkel is responding to theoretical constructs of normatively guided actions, structural constraints and normative functions of organisations in society. He is responding to certain versions of Durkheim (Garfinkel, 1967, p.vii), that sociologists should assume the objective reality of *social facts* such as social action and social order, and treat this reality as the fundamental principle in sociological analysis and explanation (Durkheim, 1985[1938]; also eg. Garfinkel, 2002). He is also responding to the application of scientific rationale and the use of study techniques inherited from the other sciences: hypothesis, evidence and inference, causal relationships and causal adequacy, rational plausibility, and so on. We shall then review some of the concepts introduced by sociology's founding fathers and look at some of the developments within sociology that draw on them.

Sociological reasoning

Both Durkheim and Weber promoted the application of scientifically oriented methods to investigate the functions of social systems and social action. For example, Weber considered generalised theoretical categories as essential to the proof of causal relationships in the human and cultural domains (Weber, 1947). Thus, Weber states in his opening of "The definitions of sociology and of social action":

> 1. Sociology (in the sense in which this highly ambiguous word is used here) is a science which attempts the interpretive understanding of social action in order thereby to arrive at a causal explanation of its course and effects. In 'action' is included all human behaviour when and in so far as the acting individual attaches a subjective meaning to it. [...] Action is social in so far as, by virtue of the subjective meaning attached to it by the acting individual (or individuals), it takes account of the behaviour of others and is thereby oriented in its course. (1947, p. 88).

What was to distinguish systematic sociology from the other sciences, Weber suggests, is the investigation of subjective categories by which the meanings of persons, things, ideas, normative patterns and motives as well as empathic interpretations of action would all contribute to adequate clarity in the sociologist's analysis and explanation.

Durkheim, on the other hand, looked to social structures and the institutions of society to distinguish sociology from psychology as an independent science. The nature and attributes of the *conscience collective*, he argued, differ from the nature and attributes of individual consciousness. Sociology as a science should be defined by the subject matters it studies, ie., `social things' such as culture, law, customs, religion, values, morality and so on. In short, Durkheim states that the properties of collectives are real *things* and that their nature asserts itself in the same way as any other natural phenomena studied by scientific disciplines:

> We arrive, therefore, at the following principle: *The determining cause of a social fact must be sought among antecedent social facts and not among states of individual consciousness*. [...] The function of a social fact can only be social, that is to say, it consists in the production of socially useful effects. [...] We can thus complete the preceding proposition by saying: *The function of a social fact must always be sought in its relation to some social end*. (Durkheim, 1972, p74).

According to Durkheim, the ways in which the objectivity of social facts is given in nature, is external to the observable behaviours of social institutions and other collectives. The sociologist is a scientific observer and sociology is a science like the natural sciences in the sense that observables can be measured against objective universal facts. But, for Durkheim, these facts are predominantly moral facts. As Giddens puts it, “he always conceived of his contributions to sociology as being primarily focussed within the more specialised field of the `sociology of moral facts'.” (Giddens, 1972).

As regards the influence of Weber and Durkheim in Parsons' work, Parsons continues to work on Durkheim's question on how societies maintain their stability over time (Durkheim, 1972), ie., by attempting to explain the internal stability and cohesion of social institutions as constituent working parts of society. Parsons also placed an emphasis on Weber's *normative validity* to position his own voluntaristic theory of action on a continuum with Weber's thesis

about the normative orientation of social actions (Parsons, 1937). But Parsons theory was deterministic rather than voluntaristic with respect to social action. As Giddens puts it, he treated “voluntarism as equivalent to the internalization of values in personality, thereby attempting to relate motivation to the *consensus universel* upon which social solidarity is held to depend.” (Giddens, 1995, p. 235).

Weber himself did not focus exclusively on purposively rational and normatively guided actions, although he emphasised the distinction between rational and irrational actions and beliefs. He warns us that,

> [i]t is naturally not legitimate to interpret this procedure as involving a 'rationalistic bias' of sociology, but only as a methodological device. It certainly does not involve a belief in the actual predominance of rational elements in human life, ... (Weber, 1947, p.92).

In other words, Weber's teachings are that the rational-irrational distinction is a methodological device. He positions the sociologist as a scientific observer, as Durkheim does, armed with objective measures. As Lynch describes it, Weber's action theory, “proposes a descriptive method that uses an idealized “standpoint” of an omniscient scientific observer.” (Lynch, 1993, p. 8). Similarly, Lynch says of Parsons,

> Parsons never let go of the theoretical leverage provided by an *idealization* of a scientific observer's knowledge of the conditions and choices relevant to actual situations. Parsons used the *imaginable possibility* of obtaining verifiable empirical knowledge of the situation as a standard for defining and distinguishing among the various subjective elements and normative standards composing a conceptual framework for action. (1993, p. 9. original emphasis).

Parsons' theoretical approach to social structures was criticised for implicitly justifying social institutions (eg. Merton et al, 1959). He was criticised for focusing on the functional aspects of institutions and ignoring the many dysfunctions that destabilise the social order. This inconsistency, highlighted by Merton and his followers, was addressed in attempts to expand on functionalism with a more flexible theoretical approach to empirical studies of the social world.

Although functionalist theories and the constructs based on them are no longer

predominant, sociological studies are still approached using social scientific methods and analyses which continue to be formulated with scientifically inspired language—the logics of principles and causal relations and a comfortable, if not idealised, standpoint for the scientific observer. Sociologists construct conceptual and topical maps with which they can apply the fundamental principle about the reality of social facts. The central theoretical challenge is still the one identified by Durkheim and Weber—how to work with theoretical dichotomies of norms and deviations, subjects and objects or macro vs. micro level investigations that oppose structure and agency and distinguish man in general from man in particular (eg. discussion in Latour, 2005).

Scientifically oriented technology and sociological reasoning

Technology appears as a subject in early modern sociology. Perceptions of scientific and technological advances are reflected upon to make social, political and historical claims. Weber's historical perception of rationalisation and scientifically oriented technology served to substantiate his `rationalistic' study method. The central concept of his method, *pure type of rational action*, would serve the sociologist as an *ideal type* from which all deviations or irrationally motivated actions can be distinguished (Weber, 1947; Weber, 1948). In reflecting upon technological advances, Weber underscored what he saw as rational progressions towards specialisation and calculability. He used the metaphors of a *political machine* and a *bureaucratic machine* and he underscored the essential elements of bureaucratic techniques, i.e., the use of documents and files. Although he stated clearly his concerns about the development of such apparatuses toward depersonalisation, he favoured what he called the `intellectualist rationalisation' created by scientifically oriented technology. This rationalisation was the logical foundation of modern bureaucracy, modern production of goods and successful democracy. Technological activities proceeded smoothly, in his view, according to their own logic towards stability, predictability and calculability as required by the modern capitalist enterprise. Accordingly, he concluded, law, economics, politics and

bureaucracy can be as calculable as a machine.

Karl Marx also refers to a historical process when he, with Engels, writes about the contradiction between rational advances of technology in production and the irrational workings of capitalism (Marx and Engels, 1998b; Marx and Engels, 1998a). Marx uses a historical method however, which was criticised by Weber for not considering generalised theoretical categories as essential to the proof of causal relationships in the human and cultural domains. Marx's materialist conception of history was based on traditions in German thought on topics such as the economy, politics, power and the social order. Thus, he spoke of the *locomotives of history*. He located, historically, the economic classes which, along with social, political and material forces, determine the means of production with class struggle at the centre as the force of change (discussion in Gerth and Mills, 1948; also Marx and Engels, 1998b; Marx and Engels, 1998a).

What the teachings of these early modern sociologists have in common is that they depict technological positivism and determinism. Marx voiced his concerns that technology determined what was happening to human labour. Technological advances were cited by Weber to support causal explanations of the course and effects of normative or habitual actions—rational and irrational. Also, Durkheim believed that ethical and social structures were endangered by technological progressions (Durkheim, 1984[1893]).

Technological positivism and determinism has been particularly resilient in the scholarly literature and so has the distinction between types of action and activities, as rational or irrational, scientific or common sense. On the one hand, there is the human life world shaped by social, cultural and historical factors which inevitably lead to some irrational actions and beliefs. On the other hand, there are scientific and technological advances. Their constitution in rationality and scientific method transcends the common life world. There are two significant manifestations of this distinction which are of interest to this thesis. One is how sociology has borrowed from science the staging of verifiable accuracy and objective explanation to solve the *problem of social order* and explain *social action* by way of analyses

which are formal, constructive, causal and interpretive, as I have outlined above. The other manifestation is how the so-called *epistemological contents* of science, engineering and technical work have largely remained beyond the reach of sociological analysis.

Technology appears on the methodological and theoretical agenda throughout modernity. It is studied and referenced in order to clarify the distinctions between science and engineering rationale, and the logics of sociological reasoning (eg. Merton, 1957). For example, Merton and his followers took a Weberian interest in the work of engineers and technologists. They looked at the influence that work had on the development of method in production and bureaucracy. They took note of the material resources and the aims and uses of technology whereby *positive* knowledge, technological ideals and economic and political beliefs were cast as having interdependent relationships in shaping the social order. They were interested in how the work of technologists and engineers is organised on the basis of social and economic decisions with social and economic consequences, and how *ideas* themselves had a role in directing these outcomes (see discussion in Barber, 1959). But studies of technology in the Mertonian tradition distinguish quite clearly between internal and external explanations. Developments that were seen as internal to a technical domain would not be explained by social-historical circumstances or origins.

This internal domain continues to be ignored in contemporary studies. For example, the technologies of what is commonly referred to as the post-industrial era, most notably computers and computer networks, have given rise to investigations into the impact these new technologies are seen to have on public, political and private lives. Giddens speaks of the role of computerised communications technologies in societal developments toward globalisation (Giddens, 2000). He describes a structural connection between the spread of a global information society and the expansion of parliamentary democracy across the world—`structural connection', according to his theory of structuration (Giddens, 1986), refers to the sets of rules and resources available to individual actors in the practices that reproduce social systems. What can be broadly described as internet or online research also seeks to

foreground how the new information and communications technologies are penetrating every aspect of our lives. Studies are conducted on a range of topics such as virtual communication, community-building and democracy (eg. Markham, 1998; Rheingold, 2000; Woolgar, 2002; Shane, 2004), information sharing and privacy (eg. Solove, 2004; Wright et al, 2008), crime and regulation in cyberspaces (eg. Williams, 2006), and on methodological considerations for online ethnography (eg. Jones, 1999; Hine, 2000).

According to the above then, sociologists have sought new avenues for conventional problem-finding and explanation in a world that is described as increasingly reliant on advancing technologies and dominated, if not determined, by technocracy. Technological progressions are typically studied in terms of impact, acceptability, threats, empowerment, rewards, social-institutional relevance, law and morality or the ongoing restructuring of a more or less democratic world. What has been at stake in these studies is the legitimacy and verity of sociological method and reasoning about a *social world*—claims which are made by way of enquiry and reasoning that does not make adequate reference to the ways in which particular goings-on are actually lived and enacted, as ethnomethodologists would put it. So, how exactly do studies in ethnomethodology differ?

Studies in ethnomethodology: respecifications

The first thing to acknowledge about studies in ethnomethodology is how ethomethodologists respecify sociology. This section will focus on what is meant by *respecification* and, in doing so, clarify what exactly it is that ethnomethodologists seek to respecify.

When Garfinkel talks about respecifying foundational matters, he is referring to any of the *social facts* that are theorised in conventional sociology (eg. Button, 1991b). The reference point for respecification is that such foundational matters are of particular interest as fundamental phenomena to be studied in their local and endogenous production, natural organisation, accountable, ongoing practical achievements—as always, everywhere, entirely

member's work (Garfinkel, 1991; also discussion eg. Lynch, 1993). In other words, studies in ethnomethodology have a *procedural* emphasis, meaning that it is only *in* and *through* step-by-step procedures that so-called *social facts* are achieved. They are *worked* or *laboured* by people and step-by-step procedures should thus be of considerable interest (Garfinkel, 1996).

With this directive, ethnomethodologists attempt to recapture the singular just-what-that-is in the details of people's labour or anything *labour* is a proxy for. Just-what-that-is, in the situated encounters between people and with the objects of their activities, achieves meaning and order, displays rationality in recognisable ways, makes histories and futures relevant, and so on. Ethnomethodology's respecification is not formulated or enacted, however, as an alternative theory or method that can be recognised as another contribution to sociology on the basis of the familiar disciplinary foundations such as the ones discussed in the previous two sections. To enquire into ethno-methods - people's methods - requires a sensibility for *what* and *how* to study which in all likelihood will fall short of being proper sociological topics or even considered as sociological investigations. For example, when ethnomethodologists enquire into what constitutes a technically-informed practice and technical work in organisations, matters such as order, action, rationality, skill and accountability are made visible in the very details of their respective achievements and thus are also respecified as occurrences of phenomena that can be studied. In a similar vain, ethnomethodologists have suggested that rather than accepting the typical preconditions for empirical sociological investigations, matters such as the nature of data, inference and evidence, epistemology and the role of theory should be made visible in the very details of their respective achievements and thus respecified as occurrences of phenomena that can be studied, rather than simply applied as would be the conventional way forward (Sharrock and Anderson, 1987; see also Button, 1991b: ch.5-6). In other words, enacting a respecification is to embed that respecification *within* descriptions of detailed, embodied action and interaction.

This meticulous resistance to theory is profoundly counter-intuitive, given the ways in which sociologists are normally trained. Ethnomethodologists come across as utterly useless

sociologists. As Button explains, they cannot say what they mean, for instance, by action, order or rationality, sociologically speaking. Rather, they speak of such phenomena by reference to the details of a particular life-world scenario where action, order or rationality is practically achieved and ordinarily intelligible as an action, an order or something rational (Button, 1991a). Ethnomethodologists also ask how it is that sociologists contemplate the nature of these matters without adequate reference to the phenomena the selfsame matters are meant to have a grasp of. Ethnomethodology's programme offers no escape from the circumstantial detail of what Garfinkel calls `reflexively accountable action', whereby the solution to the *problem of social order* should be treated as completely internal to specific self-organising sites and circumstances (eg. Garfinkel, 1967; Garfinkel and Sacks, 1970; Garfinkel, 1988; Garfinkel, 1991; Garfinkel, 1996). Thus, Garfinkel announces a respecification of sociology with respect to visible but typically unnoticed phenomena. The distinctive technical craft of ethnomethodology is to uncover these circumstantial details which are, ethnomethodologists argue, furnished with information about how structure, action, order, rationality, culture, values, histories, method, theory, etc., are achieved on an ongoing basis.

Structure and agency

To continue this account of how it is that studies in ethnomethodology are so radially different from conventional sociological reasoning, I will take an example of the traditional structure versus agency debate in social theory. Studies in ethnomethodology are often characterised as agent-centred, micro-level investigations – or in any case a variant thereof (eg. Giddens and Turner, 1987) – so my reason for choosing this example is to elaborate how this is not the case.

Lévi-Strauss, Althusser and Parsons, qualify as key leaders of the 20th century in arguing against the self-determining capabilities of individuals (Lévi-Strauss, 1963; Althusser, 1971; Parsons, 1937). Large-scale organisations across society produce determinants of conduct of

which the individual actors are unaware for the most part. The opposite position, gaining momentum in the late 1950s and 60s, argues not to underestimate the contributions of individuals to the nature of social reality (eg. Goffman, 1956; Becker, 1963). Individuals participate in social functions that establish systems of reciprocal and symbolic interaction, establish and confirm meanings and norms, label members as of this or that membership, and so on (see also Merton et al, 1959). There have been attempts to reconcile or, say, ease the implicit tensions between these opposing views on the reality of the social order. Thus Giddens, for example, deliberately avoids privileging one over the other of the structure-agency or system-actor pairs (eg. Giddens, 1986). His analysis underscores that individual actors can draw upon pre-existing sets of rules and resources which are contextually resident within an existing structure in a particular domain. He argues that moving influences in our practices reproduce social systems and that sociological research can identify new sets of influences. The methodological implications are firmly rooted in sociological traditions and the goal is to reveal the contextual rules and resources with which action occurs, and thus also the norms and laws that render action *partially* predetermined. From this and similar positions on the matter, we have different levels of social reality, different degrees of internal and external, local and global causes (see discussion in Latour, 2005).

Some of the radical shifts from structure-centred to agent-centred analyses relate to specific study topics such as the studies of organisations and the technologies of production. For example, in his Theory of Organisations, Silverman argues against the idea that technology is operating on an intervening environment variable with other structural constraints that can influence behaviour, interactions and sentiments in predictable ways (Silverman, 1970). His argument runs against technological positivism in systems approaches to organisations, ie., viewing behaviour, motivations or interactions as an outcome of the technologies of production and the structures within which they operate. Silverman criticises the positivists for seeing a mechanical relationship and taking certain human needs as universally given, such as the need for meaningful work group relations or the dissatisfaction

with repetitive work on assembly lines (eg. Walker and Guest, 1952; Sayles, 1958; and Blauner, 1964; cf. Silverman, 1970: ch.5).

In casting this critique of technological positivism and systems analysis, Silverman filters his own conception of technology into a new theory of organisations. Behaviour is not predictable from knowledge of technology and human needs are not universally given (see also Goldthorpe, 1966; Reeves, 1967). Rather, people establish meaningful and shared understandings of their activities in organisational and technological settings. Action arises from meaning which is both locally established and culturally inherited, in Silverman's view. People will react to technologies and shape their activities and sentiments with regard to technological functions and organisational structures in ways that are not necessarily predictable.

There is no doubt of a radical shift in Silverman's thesis from previously dominant views. Firstly, his thesis is a flat out rejection of technological determinism which, one can argue, coincides with the trend of the day. Secondly, he suggests that the nature of organisational life is radically different from what previous studies argue. He makes a point of addressing technologies of production as a significant part of that life and then argues that actions of individuals not only sustain the reality of the social order, they change that reality.

Silverman's main focus in this thesis is theoretical—to place *meanings*, from which action arises, in the context of the logic of an established academic discipline. His influences are, among others, Goffman, Goldthorpe, Glaser and Strauss (Goffman, 1956; Goldthorpe, 1966; Glaser and Strauss, 1968), and he strongly promotes a social action model and grounded analysis which is continuous with interpretive sociology in the Weberian tradition, as he states himself (Silverman, 1970, p.222-223). Sociologists, he argues, can attain the subjective understanding of actions and generalise on the basis of a probability of actions in terms of typical motives or intentions. Typical acts of typical individuals are studied with the use of ideal-types that take account of subjective meanings and choice, oriented to other actors and their intentions. But, as he later admits, his theory is only vaguely suggestive of

how to apply it (Silverman, 1994). It prescribes a theoretical goal—to explain the generalised assumptions upon which meaning arises among actors, upon which work and organisational life are thought to be based.

At a later date, Silverman expresses a number of reservations about his Theory of Organisations (eg. Silverman, 1994; Silverman, 1997). One of his concerns is that his strong reaction to structural constraints might have pushed an anti-structural and individualist agenda, relying too heavily on states of mind. Another concern he raises has to do with the vagueness of how to apply his theory. In discussing the matter, he points to studies in ethnomethodology as an example of a technical craft that is superior to his social action theory, in his view, in that it uncovers how actors orient to relevancy constraints and how they accomplish and make observable-reportable their everyday reality. He does not abandon constructive interpretive analysis, however. Studies of folk methods serve his purpose as one piece, albeit an important one, in a constructive analytic approach to the study of organisational practices which draw on conventional perceptions of structural constraints and historical origins. So how do ethnomethodologists respond to the suggestion that studies in ethnomethodology are agent-centred, micro-level investigations that can be reconciled with sociological traditions as Silverman indicates?

Ethnomethodology and the social actor

In a piece on how studies in ethnomethodology respecify the social actor, Sharrock and Button address the traditional *structure* versus *agency* opponents. They admit that, “it may seem that ethnomethodology is best understood as falling squarely on the `agency' side.” (Sharrock and Button, 1991, p. 138). As Lynch has also remarked, the term `ethnomethodology' has been used “to describe any of a variety of ethnographic or hermeneutic approaches to situated social practices” (Lynch, 1993, p. xvii). And, he continues, “[c]onversation analysts have advanced increasingly formalist and foundationalist claims about language use” (Lynch, 1993, p. xviii), whereby the ethnomethodology's

programme, aka Garfinkel, serves at best as a historical curiosity—a more or less forgotten forefather.

Studies in ethnomethodology and offshoots like conversation analysis may not be uniform in character nor in emphasis on what and how to study. But the main considerations that are taken into account in this thesis follow directives from Garfinkel himself, as I understand them, and particularly relevant are studies conducted by his followers whose focus has been to uncover the details of lived technical work and technologically-assisted practices. I will come to that shortly. What Sharrock and Button explain is that in order to belong on the side of agent-centred sociological analysis, ethnomethodologists would have to engage in debates on such matters, as outlined in the previous section. They are pitched at an ontological level about the reality and nature of fundamental entities in sociological investigations such as action, structure, agency or order. As Sharrock and Button remind us, there is no need for these theoretical constructs in studies of ethnomethodology. There is no work for them to do. When such theoretical constructs are put to work by sociologists they can theorise the *problem of social order*, for instance, or any of the topics that *order* is a proxy for (Garfinkel, 1988). Sociologists need these constructs to hypothesise stable external arrangements and describe what is characteristic of persons and relations:

> ...external to the orderliness observable in the sites of everyday activity. Ethnomethodology's respecification is, however, to treat the solution to the 'problem of social order' as completely *internal* to *those sites*. It conceives social settings as *self-organising* and for just that reason has no further need for the received concepts... (Sharrock and Button, 1991, p. 141 [original emphasis]).

What I am emphasising here is that studies in ethnomethodology are resolutely unconstructive, at least from any theoretical standpoint from which they were dislocated by Garfinkel himself. These studies illustrate how the self-organisation of social settings is achieved. The *problem of social order* is not an inductive problem which needs to be addressed with theoretical generalities to cover a sum of activities, incidents and circumstance. Ethnomethodologists give priority to the actor-in-a-social-order, the actor's

point of view and the irreducibility of events-in-a-social-order. But that is not the same as committing to a theoretical position which argues that a *social reality*, objectively speaking, can be reduced to an ensemble of interactions, activities and points of view. According to Garfinkel and his followers, theoretically informed sociology remains fundamentally flawed, namely, sociologists rely on underlying or, essentially, *external* social variables to theorise action, events and the social order (Sharrock and Button, 1991). As Sharrock and Button point out, studies in ethnomethodology are actually responding to the problem of getting theories to attend to that which is plainly before one's eyes by attempting a recovery of primitive or *pre-theoretical* phenomena—that which *can* be seen but typically goes unnoticed.

I will now turn to a review of investigations into the aforementioned *epistemological contents* of technology and foreground the sociological claims that are made about them through problem-finding, empirical investigations and theoretical constructs. In doing so, I will also clarify how studies in ethnomethodology respond to these so-called `critical investigations' and how they differ by emphasising a recovery of pre-theoretical phenomena.

Technology studies

Investigations into the internals of technology belong to a relatively small division of the social sciences. An interest in the *epistemological contents* as a legitimate topic for sociological analysis followed a significant shift in thinking about scientific knowledge claims (Barnes, 1974; Bloor, 1976; Collins, 1981; Collins, 1983). Sociologists of scientific knowledge (SSK) were breaking with the rational-irrational distinction of how and why we come to believe in facts or false claims. They were taking new steps by arguing that *all* knowledge claims (true or false, scientific or common sense) are socially constructed strictly speaking or, in any event, thoroughly socially constituted. Philosophers and historiographers also paved the way for sociological interest in technical work (Wittgenstein, 1953; Polanyi, 1962; Kuhn, 1962; Feyerabend, 1975). Although the works by these scholars are diverse in

approach and the claims they make, they provided sociologists with the means to hold the `social' responsible for conditions, decisions and concerns of technologists and technical labourers (see discussion in Lynch, 1982). What followed was empirical research to show how technical work and artefacts are socially constructed.

The study programmes that developed, and are now well established in Science and Technology Studies (STS), were attempts to break open the `black boxes' of technology. Historiography and sociological reasoning was put on the agenda, aimed at explaining technical artefacts, systems and technical work on the basis of social-historical origins, circumstances and social epistemology. Discoveries of positive knowledge and technological rationale were no longer accepted as intrinsically rational and positively ascertained. The internal-external distinction had been abandoned and so had technological determinism. Inventors were no longer central explanatory subjects, carrying the weight of a journey from discovery and invention to innovation.

Opening up black boxes became, as Jordan and Lynch put it, a *rallying cry* to contextualise the technical by revealing hidden motives, interests and origins in a social domain (Jordan and Lynch, 1992; see also Lynch, 1982; Lynch, 1985; Lynch, 1993, pp.265-271). Lynch argues that the very `criticalness' of these efforts has been achieved by portraying practitioners and technical labourers as if they lack in ability to access the circumstances of their own work—or, what is *really* going on in the decisions they make: “The technician is unaware of the material conditions of his knowledge in an alternative domain—the social context of his practices” (Lynch, 1982, p.510). This conception of inadequacy has been instrumental in, what Lynch calls, a disengaged application of sociological reasoning and historical methods that specify what the origins and circumstances are. We shall now examine some examples of these developments within STS.

Pinch and Bijker's thesis on the social construction of technical artefacts is presented in a story about the early developments of the bicycle (Pinch and Bijker, 1984). As Pinch and Bijker put it, the story is focused on multi-directional progressions in the development of

artefacts. It reveals that there were many options in the bicycle's design: higher and lower wheels, differently shaped frames, the position of the seat, the option of an air tire, so on and so forth. A number of bicycle types came and went and to make sense of this development, they argue, there are social variables to consider. *Relevant social groups* have to be identified. Social groups have *high or low inclusion* in the development of the artefact. They have different *stakes* and *interests* as potential user groups (e.g. men, women, young, old or sportsmen) or as engineers, available suppliers and investors. The *interpretive flexibility* of artefacts is also critical, meaning, that members of different groups interpret artefacts and their functions differently. These interpretations and potential selection in design are *negotiated*. Finally, the *framing* and *re-framing of problems* such as issues of comfort, safety, speed and clothing, will influence whether individual solutions along the way are seen as successful or not, whether or not they include or exclude user groups, and so on.

The basic conception of how to explain technical artefacts and practices, in this and similar studies, is constitutive of a new frontier for sociological problem-finding and argumentation (eg. Bijker, 1987; Schwartz Cowan, 1987; Elzen, 1986; MacKenzie, 1990; Misa, 1992). These studies essentially echo conventional sociology by asking, as Merton put it, *why* the outcomes have turned out as they did (Merton, 1959). What we are learning new however, is to approach the study of technology on the basis of the methodological relativism introduced by sociologists of scientific knowledge (Bloor, 1976; Collins, 1983). Investigators were seeking to attain an interpretive understanding of technology among the persons/groups they were studying, and also to arrive at an interpretative understanding of the causal explanations for how the contents of variable interpretation and negotiation are socially constructed. What the data uncover when the black box is pried open, they say, are social processes on the bases of which both failures and successes can be explained. In other words, both successful and failed technologies are socially constructed and socially determined rather than determined by an *objective technological rationale*.

If we follow the developments in technology studies, we also learn that this new frontier

has divided STS scholars on a number of issues. One concerns the privileging of strictly social or social-historical explanations of *why* things turn out as they do or *why* practices proceed as they do (eg. Law, 1991; Woolgar, 1991b; Bijker and Law, 1992). Another concerns the use of analytic categories such as social vs. technical and human vs. object, and conventional disciplinary and other group distinctions, ie., whether they represent distinctly technical, managerial, economic, social, political, or end-user involvements (eg. Callon, 1987). As Latour has argued, the boundaries are too fuzzy when studying technologies and organisations (Latour, 2005; see also Latour, 1991; Latour, 1993). Things change too fast for the ingredients to be comfortably known.

Studies of use and user trials are instructive in this respect. For example, drawing on Woolgar's concept of the *configuration of the user* (Woolgar, 1991a), technology users are depicted as a fragmented group of visionaries, developers, administrators, lawyers, technical operators, end-users (men, women) or others, who all relate to, interact with, and inform a particular technology at one stage or another (eg. Mackay et al, 2000; Oudshoorn and Pinch, 2003; Oudshoorn et al, 2004). People also (re)configure artefacts and systems in unpredictable ways. For example, Akrich investigates the practical rationale of *de-scripting* and *re-scripting* devices (eg. Akrich, 1992) and, drawing on her investigations, considerable attention has been directed at *workarounds* both in design and use (eg. Garrety and Badham, 2004; Pollock, 2005). Features are ignored, artefacts are modified, and so on. Thus, we hear of messy contingencies of social and material resources (eg. Latour, 1988; Latour, 1991; Latour, 1993), as well as of seamless webs of technical, economic, social and political considerations (eg. Hughes, 1985; Hughes, 1986; Callon, 1987). In other words, technology studies have found an indomitable fusion of disciplines and roles in the development and deployment of technologies.

By identifying interests and stakes and the interpretive flexibility in negotiations and closure, investigators are explaining why their subjects come to agreements or beliefs about material phenomena. But technology studies also moved on to study technologies in-use, that

is, to say, in production, control, support or other roles, each of which introduce complications with respect to the very concept of a black box as well as the project of prying it open. For example, studies of technologically informed practices in organisations have sought to capture and theorise the complexity of technical systems and of the organisations themselves as complex evolving socio-technical systems (eg. Gerson and Star, 1986; Star and Griesemer, 1989; Bijker and Law, 1992; Law and Mol, 2002b). According to Gerson and Star there are gaps between multiple viewpoints in organisations—how work is articulated, planned, realised, represented and made accountable (Gerson and Star, 1986). There are perpetual inconsistencies among evolving bases of knowledge and skill, they argue, and work can be characterised by ongoing negotiations about social and material relationships, by ad hoc reactions to contingencies that arise with respect to material constraints, uncertainties about how to solve particular problems, so on and so forth. To use Gerson and Star's way of putting it, we are looking at open systems that must cope with multiple competing, and possibly irreconcilable local solutions. In the field of organisation studies, more specifically, both social action theory and decision-making theory have been criticised for failing to deliver satisfactory results insofar as being instrumental in improving understanding and administration of organisational life (Silverman, 1997). One alternative has been to demonstrate the value of metaphors (Morgan, 1986; see also discussion in Silverman, 1997). Organisations have been variously cast as instruments of domination, as cultures, organisms, brains, systems, machines, etc. Morgan's argument is that "[i]mages and metaphors are not just interpretive constructs used in the task of analysis. They are central to the process of *imaginization* through which people enact or `write' the character of organizational life" (Morgan, 1986, cf Silverman, 1997, p. 180 [original emphasis]). The acts of imagining and actually `organising' organisational life have also been variously described as ordering materials, people, categories or technical systems (discussion in Law and Mol, 2002a). Law and Mol suggest that each metaphor or description of these activities captures some version of conditions and affairs but different versions can also co-exist independently, however

much they overlap and interfere with one another (see also Knorr-Cetina, 1999).

According to these conceptions, technical artefacts – their development, interactions with them and the understandings of them – are tried and configured in ways that do not necessarily lead to closure except in the sense of achieving *end products*, *organisational configurations* or *practices* with uncertain futures. What shall we then make of the concept of the black-box and how is it helpful (or perhaps unhelpful) to technology studies? To put it crudely, whether the concept of the black box stands for the manifest outcome of strictly social processes or the outcome of evolving socio-technical practices, then the concept can be used to characterise any well established fact, artefact or practice. It can also be assumed then that uncertain facts, artefacts and practices are potential or *not yet* black boxes. And, if the assumption is that black boxes are indeed the eventual outcomes of social or socio-material histories (presumably often contentious ones), then it is worth asking to whom are they opaque or obscure as opposed to transparent or translucent and open to further innovation and development. The claim that *closure* is what investigators are finding the processes of, as described by Pinch and Bijker (1984) or that they are finding forgotten or overlooked histories, is far from being unproblematic (see also discussion in Jordan and Lynch, 1992). A range of studies within the STS tradition indeed suggest that *uncertain* facts, artefacts and practices are integral to everyday goings-on in the operations of complex systems, in laboratory work and more. One could ask then, with good reason, whether such uncertainties necessarily demand closure or whether they are incorporated and dealt with in some other way.

Regardless of method or detail in describing technical practices, artefacts or systems, most of the studies mentioned in this section address the so-called *epistemological contents* of technology. The interpretive constructs are technical concerns, decision-making, problem-solving, the ordering of objects and systems, as well as the social, material or historical origins of why things are as they are. What these studies also have in common are contemplations about the nature of sociality, organisations, rationality, technical work,

technologies-in-use, and so on. What is of concern here is not if one method or theory is more plausible than another in explaining what goes on, sociologically. Rather, it is the question typically set aside—how are technologies and technical practices *enlivened* both privately and collectively?

Ethnomethodology and STS

What ethnomethodologists have alluded to in their contributions to technology studies are only indications of what can be uncovered if actual occurrences of interactions and work were adequately regarded. In this section, I will take two examples, both of which look at technical work in a laboratory. The first one examines the visible-reportable enquiries into social-historical origins and circumstance in the course of doing technical work. The second one challenges the concept of the black-box and the project of prying it open in relation to uncertainties in particular routine procedures.

Lynch argues that laboratory scientists speak and act in each others' presence in ways that exhibit richness and specificity to questions of the contingent ties between their objectified results and the social-historical circumstances of their work (Lynch, 1982; Lynch, 1985). They perform distinctive modes of critical enquiry, he argues, which are necessarily a part of what they normally do. For example, they have to commit to questions of origins in doing and communicating their work: “in order to demonstrate that they are achieving results with `responsible' origins” (Lynch, 1982, p.512). So, their technical practice is enlivened with troublesome open-ended questions of finding in the visibility of a result that a procedure was well constructed or that something is amiss and an unforeseen contingency cropped up in the local history of the particular task. When observed in spontaneously produced interactions between the practitioners, Lynch argues, these endogenous occurrences of enquiry also have generative claims upon further progressions of the work.

In the 1982 paper, Lynch offers the reader 34 lines of transcribed conversation. Laboratory practitioners are remarking upon an electron micrographic montage, produced as

part of an ongoing project. Claims are made about what is seen in the montage, how the mircographs were shot, what else is involved in doing this work, and if or how this particular product is useful to the project. Lynch explicitly avoids classifying circumstantial matters, "involving `social', `personal', `instrumental', or `objective' domains of explanation." (p.512). Rather, as Lynch takes his readers through the transcript, line by line, we learn that matters of *social relevanc*e are those of "practical responsibility, competence, and fortune — manifested in the concrete appearance of `degeneration', `starting points', and whatever else is pointedly named" by the practitioners (p.518). The references – the named phenomena – continually point to or, as Lynch says, they *attach* themselves to an object-at-hand and the manner of that *attachment* is, for Lynch, specifically at issue over the course of the sequence.

> These references signify fields of observables [...] They include, as well, the practical and personal sources of the object's presence 'here' — its perspectival features 'oriented' to the embodied work of 'reading' it, its 'unfortunate' and 'bad' characteristics, and its prospective usability. (Lynch, 1982, p.518).

In explaining to his readers what all this has to do with critical enquiries, Lynch takes as example that *hidden origins* of this visible documentary product of the work, the montage, are actually integral to the very course of a project the practitioners are working on there and then. He explains how this is made available:

1. The parties in the sequence above start out with an object-in-hand, though what the object turns out to be is interactionally negotiated.
2. The object is referred to as a definite product of a practical horizon of work with 'unfortunate' features.
3. The horizon of work for which the object stands as an 'unfortunate' product is explicated as an indefinite history with disputable characteristics.
4. Issues of fortune and personal accountability are worked out in reference to the features of the electron micrographs at hand. The micrographs are treated as the visible ground from which assertions and agreements on matters of authorship, competence, and responsibility are explicated. (p.519)

The transcribed interaction is omitted here, but the first thing to notice about 34 lines of exchange is how insignificant they seem at a glance. Step-by-step reading through the account however, uncovers arguments and claims. They are *localised*, as Lynch puts it, in the immediately pertinent details of an object-in-hand, which is also pictured in Lynch's paper.

What is different about the enquiries of sociologists and historiographers, compared to this example, is that we learn that technical work has a history *behind* itself, complex social or socio-technical circumstances *surround* it, and it progresses *through* a historical terrain. What Lynch is offering is not a negation of histories, circumstances, progressions and terrains. What he makes available is access to technical work as an accomplishment that involves ongoing re-examination and articulation of histories and circumstances, so the progression of the work "from within involves an articulation, for all practical purposes, of what that socio-historical terrain consists of as immediately pertinent details". (p.519).

A similar argument is developed in another study of technical work in laboratories when Jordan and Lynch set out to examine how practitioners specify social and historical conditions in their work (Jordan and Lynch, 1992). This study is based on accounts by various practitioners on what is involved in purifying and isolating plasmids for molecular biology research. The accounts are supplemented by uncovering significant details in carrying out this procedure which is referred to as `the plasmid prep'. But, as I already mentioned, Jordan and Lynch make a point of casting their enquiry in reference to the interest sociologists and historiographers have had in the concept of the black box. In doing so, they enquire into the use of the concept and how the plasmid prep, as a procedure and product for a variety of research practices, is or is not such a box.

As Jordan and Lynch describe it, opening up black boxes wherever they could be found was a response to the Mertonian tradition of distinguishing between external and internal explanations—to treat scientific facts, artefacts and well established technical procedures as, precisely, black boxes (see also eg. Winner, 1993). Jordan and Lynch point out the potentially "unfortunate, or even absurd, consequences" of such a task (Jordan and Lynch, 1992, p. 111). They claim that:

> sociology's aspirations to transcend merely situated experience can lead its practitioners to assume a (hypothetical) "position" for an inquiry that is simultaneously confident in its epistemic privilege and (in O. Sacks's terms [1987]) "unimaginably remote" from common experience (p. 111).

As Jordan and Lynch remind the reader, the concept of a black box can define a range of devices, mechanisms, processes, causal laws, entities, artefacts, simplifications, innovations or palpable outcomes, all of which can then be uninteresting, in no need of further enquiry, inaccessible, glossed over, unknown, lost to history or, for all practical purposes, forgotten. They insist that the technical work of purifying and isolating plasmids has no definite status in a historicist scheme of opening up a black box. They do observe nevertheless that practitioners in molecular biology research consider the plasmid prep as practically uninteresting—an indispensable core instrument in what they do and, for all they know, a procedure that is reliable enough, written up in manuals, and so on. Observations of and interviews with those who purify and isolate plasmids however, render conditions of instability and fragmentation visible. Among the technicians, it remains contentious to what extent the procedure is established, reliable, ready-made, ritualistic, improvisational, superstitious, irrational, and so on. For example, individual practitioners talk about accomplishing habitual routines, but they do not speak in unison about *just how* ritualistic these should be or *just what* is included in these routines. They speak of using manuals, having luck, taking both rational and irrational steps, and so on. What works in the end for each of them becomes, for all practical purposes, the *rational* procedure for that person to adhere to.

To sum up, Jordan and Lynch speak of dispersion and privatisation of techniques which then lead to sporadic invasions of variants in what otherwise might be classified as the *same* procedure or the *same* product. This production of variants, they observe, is painstakingly obvious to the technicians themselves and well known by practitioners who use the product. However, they argue, such fragmentation and dispersion of techniques would be neglected in social-historical studies, especially those that are preoccupied with stabilisation and closure.

Discussion: reasoning and enquiry

The one claim consistently argued throughout this chapter is that studies in ethnomethodology break radically from conventional sociological studies. Exactly how they differ may still seem rather puzzling since little of ethnomethodology has yet been explicitly demonstrated on these pages. That ethnomethodology continues to be a counter-intuitive study programme also suggests to me that the difference between sociology and ethnomethodology is never adequately explained or demonstrated. For example, in the case of Silverman and many others, *doing* ethnomethodology has served as another resource to assess what the local achievement are or can be against the forces and constraints that lie *behind* or *underneath* them (see also Watson, 2002). STS research into expert knowledge and know-how has also shown tendencies to assimilate the writings of Garfinkel (and Wittgenstein), as Lynch points out, "into modes of sociological explanation that show little cognizance of the more radical anticausalist and antiepistemological implications of those writings." (Lynch, 1993, p.75-76). What I will do in this final section, therefore, is to retrace my steps and clarify why I find it of such importance to emphasise what this radical break actually does as a programme of enquiry and how it is that this programme is a resource for my study of the archive.

Ethnomethodology's respecification of so-called foundational matters within sociology opens the door to an altogether different set of interests—the enquiry into naturally occurring phenomena *in* and *through* which achievements such as order emerge. What this `dubious' activity has uncovered can be said to discredit the scientific status of sociology insofar as sociologists themselves argue their claims about *social facts* by theorising the functions of *social action*, *structural constraints* and *social forces*. Studies in ethnomethodology ignore common explanations of what social action and the social order consist of—explanations which are arrived at by structural analysis, inference or interpretation, using a language of causal relations and causal adequacy. But sociologists will not necessarily see that the foundations they take for granted are radically challenged, if they take any interest at all.

What perhaps troubles sociologists the most is the inability of studies in ethnomethodology to explain the *social context* and take in account *constraints*—the fact that persons behave and act either *according* to or in *deviation* from conventions, morality, rules and directives, and so on. For example, the technical workers in Lynch's 1982 study name and point at particular phenomena in fields of observables. In doing so, matters of *social relevance* emerge, ie., the practical and personal sources of the pointedly named phenomena, as well as responsibility, competence and fortune in doing this work which is an ongoing project with a history and a future. But where is the sociology in that, one might ask. Why are these goings-on not contextualised in a frame of normative reasoning about laboratory practices to which the scientists adhere, the social construction of know-how which determines whether or not members of the team know what they are doing, or the history of *this kind of research* which puts constraints on what are legitimate contributions to a particular domain of laboratory research? These factors would certainly be in play as the external social and historical forces that impinge on the local activity and affect its course. The problem with this line of argumentation however, is that it assumes a structural abstraction of insides and outsides. It assumes that forces work from remote over a terrain which is imagined both as spatial and temporal. So, if the technical workers are shown to behave in a manner that is typical of what is expected of them or happens elsewhere in similar settings, then social forces have indeed been measured and shown to be at work over space and time.

In order to make claims like that, one has to actually believe there exists such a *thing* as the social in much the same way as Durkheim proposed—the *social* whose existence was positively asserted as real in the 19th century in order to establish a scientifically oriented discipline for the study of *social thing*s. I do not subscribe to this belief. To summarise, the logics of conventional sociology hinge on the use of spatial metaphors like macro structures with inroads, externals and internals, and, most importantly, the assumption of an ontic *social*. The contents of social variables, constituted by that *thing*, are the key instruments in developing *socio-logical* explanations, whether they are based on quantitative or qualitative

approaches. But as Latour points out, this nebulous social is not the name of a glue which can explain how the world is held together (Latour, 1988; Latour, 1991; Latour, 1993; Latour, 2005). What are ethnomethodologists saying then when they argue that solutions to the `problem of social order' should be treated as completely internal to specific sites (eg. Sharrock and Button, 1991)?

When ethnomethodologists refer to the singular just-what-that-is and what just-what-that-is achieves *completely internally* to a specific site, the internal is plainly the site of reference, as I understand it—a particular site where some action and interaction takes place. It is not the same as suggesting a spatial metaphor which is exclusively focused on an internal domain, ignoring the external ones, and thereby negating the relevance of histories, institutional arrangements, various forms of policing or a sense of normality and abnormality. Rather, it is just-what-that-is in the pertinent details at hand which brings issues of *social relevance* to light and life, some of which make reference to histories, institutional arrangements, policing or a sense of normality. So, why is it counter-intuitive to argue along these lines and what exactly is my position on this matter?

The ways in which we are trained to *sort out the world* (formally and informally), and thus to explain to ourselves and others the various material, mental or social orders, hinges on the mastering of a range of sorting techniques, of constructing relationships between entities (things and persons) and detecting causes and effects over space and time. This is an acquired ability which lends itself to the assumption that orders can be objectively discovered through observation and measurement, deduction or inference, and verification against a world *out there*. It also lends itself to philosophical scepticism about verifiability—how can we be absolutely sure we (our senses, our shared or individual understandings) are not being deceived by *the world*, by others or by a devil for that matter? (eg. Descartes, 1991[1637]; see discussion in Sharrock and Anderson, 1991). I am not a philosophical sceptic in the sense that I see no mystery in the possibility of *orders*, thus achieved, serving as evidence and means for all manner of practical purposes. I see no mystery either in how the analytic

programme of ethnomethodology:

> ...builds on the “social fact” that singular instances of conduct are intuitively recognizable and vernacularly describable; [... that analyses rely] on the existential “fact” that any inquiry will find itself “thrown” into an already intelligible world, even when that world's intelligibility includes indistinct, arguable, and doubtful features. (Lynch, 1993, p.285-286).

It may seem that ethnomethodologists are playing a devious trick but they are not, as far as I understand. They are challenging particular ontological assumptions however. It is that challenge which also forms the basis for my own interest in pre-theoretical phenomena, the details of what happens internal to specific sites and how it is that in these details are *reinventions* and a *structuring* of the sort of *things* sociologists typically cast with more general descriptions and assume are largely the effects of external social circumstance. It also forms the basis of my interest in enquiry that does not, as Lynch puts it, try “to explain those details by “connecting” them to a corresponding set of contextual “factors”” (Lynch, 1993, p.30). In light of these considerations it should be mentioned as well that not all theorising projects in sociology are strictly concerned with a theory of action and the social order. In fact, the very term `theory' is often loosely applied to stand in for concepts and metaphors as heuristic devices—the uses of which can be judged on the basis of how well or poorly they do what is purposely set out with by using them in the first place. Morgan's argument in favour of metaphors is a case in point—they are “central to the process of *imaginization* through which people enact or `write' the character” of a living breathing world of people going about their business (Morgan, 1986). It is perfectly intuitive and reasonable to be imaginative and experimental. Metaphors are powerful ways of depicting what goes on and to experiment with ways to *think* out loud about certain events and activities. Similarly, organising discussions about data under themes or topics are useful instruments for composing a coherent reading experience with reference to recognisable concepts and categories. Such concepts and categories do not need to be technical, sociologically speaking. I will come back to themes, concepts and metaphors in Chapter four where I discuss my data

and what I have done with it. The next chapter will explore how I situate my study of the archive in relation to studies of work, workplaces and the use of computer supports.

Chapter 3

Computer-supported work

Computer systems and computer supports

The previous chapter discussed how studies in ethnomethodology respond to conventional social scientific and historical reasoning and, in doing so, suggest an altogether different sociology. The chapter concluded by clarifying what I expect from enquiry in relation to these differing views on what and how to study. In this chapter, I will look at studies concerned with the design and use of computer supports. As I explained in the introductory chapter, the archive comprises an amalgam of technologies *in action*. Operations, maintenance and ongoing design are achieved in virtual computer-supported workspaces and the people who work there are physically-geographically distributed. The archive is also accessed from remote by those who prepare and upload their preprints. It is therefore of considerable interest to situate my study of the archive in relation to studies of work, workplaces and the use of computer supports. I will address common considerations for the design of expert and support systems and I begin by reviewing critiques of cognitivism and similar assumptions that typically underpin these designs. I will also look at conceptions of knowledge and skills, how work is organised, how systems and supports are located in relation to human activities, and how various settings can be said to resemble the conditions for operating and using the archive.

Many of the studies I will be looking at make use of `post-cognitivist' theories, as Kaptelinin and Nardi call them (Kaptelinin and Nardi, 2006). By that they argue that critiques of cognitivist assumptions have influenced the use of alternative theories to explain what is going on in the workplace, for example, activity theory, theory of distributed cognition, social constructivism and actor-network theory. Again, I will be looking at the ways in which

studies in ethnomethodology differ from such theoretically informed studies and in what way this difference is of importance to the development of this thesis.

Critiques of cognitivism

Cognitive science and cognitivism have come under criticism in relation to research and engineering of artificial intelligence (AI), human-computer interaction (HCI), computer-supported cooperative work (CSCW) and, more generally, of the organisation of work and institutional practice. A considerable body of this critique is focused on failures in the design of so-called expert systems and of information and communication systems. Although studies of design and design failure are not of immediate interest for this thesis, the critiques on which many of these studies are based are important to consider.

Cognitivist assumptions have been blamed for failure. The human agent is explained on the basis of processes in the mind (or brain), and interactions with world are explained on the basis of `input/output' functions, information processing theories, cognitive control structures, mental models, and the like (Carroll, 2003). In other words, cognitive science is a study of isolable agents with specifiable attributes and resources, and HCI and AI researchers who draw on the science, work with formal descriptions of knowledge and tasks. But cognitivism has also been critiqued in relation to particular sociological and philosophical interests. We shall then explore some of the critiques whereby such interests are foregrounded.

Sociologists seek to explain the individual as a social agent on the basis of sociality and social processes, and interactions are understood as governed by social-cultural conventions. Sociologists (and anthropologists) are also interested in a wide range of knowledge – learned and applied – which, they argue, is not given due consideration in expert systems or in HCI designs. For example, Forsythe comes to the conclusion that the cognitivist conception of knowledge is far too narrow (Forsythe, 2001: ch.3). AI researchers, she argues, divide knowledge into expertise and common sense, and focus on the former in their attempts to elicit it from the experts for design purposes. Her claim is that essential aspects of knowing

and doing are invisible, overlooked or undervalued. Drawing on Susan Star, she comes to the conclusion that male-centred and rationalistic values dominate research and design methods, and effectively *delete the social* (see Star, 1991; also Forsythe, 1995).

Harry Collins, in his critique of expert systems, takes a similar view to Forsythe's in that only the formal aspects of knowledge can be elicited in formal ways (Collins, 1990). Collins' main argument however, is that expert knowledge is acquired in the course of being socialised into a practice or a *form of life* (see Wittgenstein, 1953; also Polanyi, 1962 on tacit knowledge). Correspondingly, how knowledge becomes tacit and a *form of life* enculturated can only be explained, in Collins' view, on the basis of social construction and social epistemology (also Winch, 1990; Bloor, 1976). Such *taken-for-granted* knowledge is impenetrable for the purposes of reconstruction in expert systems (Collins, 1992: ch.5). Computers simply are not and cannot be social and enculturated like humans.

The idea that humans and computers have social characteristics in common is generally not entertained. Rather, a conception of `social' as the sum of specifiable interactions between individual agents has given rise to the idea that computer systems can replicate such interactions. It is not an assumption about common sociality however, as sociologists would contemplate its nature, but an assumption about common cognitive characteristics. The brain is explained in terms of computer hardware and the mind in terms of software. These explanations are what the phenomenologist, Hubert Dreyfus, calls the *psychological* and *biological* assumptions in AI (Dreyfus, 1992). He has been very critical of these assumptions and his main concern is that they represent a model of awareness and knowledge of the world which is completely disembodied. According to Dreyfus, this is not a credible model and his critique of it is effectively aimed at any philosophical or scientist assumption about knowledge that trivialises the role of the body and of *being in the world*.

Anthropologists and activity theorists, who study the design and use of support systems in workplaces, have also foregrounded the relevance of tacit and embodied knowledge when computer supports are designed. Their interests cover a range of other topics however. For

example, they speak of organisational learning (Engeström, 1996; Engeström, 2000), cultures of practice (Downey, 1998; Forsythe, 2001: ch.6-7) and they are interested in the social processes *surrounding* or *behind* the changes that occur when computer technologies are introduced into organisations (eg. Kaptelinin and Nardi, 2006). More specifically, they seek to foreground what is understood to be the working intelligence and intentions of people who work together, what motivates them and how their work can be transformed when new support systems are introduced.

When these scholarly interests are taken together, the main conclusions are that designs of computer systems and supports should be sensitive to people's motivations and intentions, their interests, culture, skills and learning processes, bodies and the *tacit dimension.* Many of these same conclusions were also adopted in organisation studies. For example, decision-making theory was the alternative to Parsonian functionalism in the 1960s (e.g. Gyert and March, 1963). It supported studies of social action and interaction, however, strongly committed to rationality as a transcendental value (see discussion in Silverman, 1997). This so-called `cognitive school' in organisation studies has since been critiqued by those who shifted their interests away from decision-making theory and behavioural-economic models, toward a sociological model of institutions and work—study agendas more akin to social constructivism and interpretive sociology or anthropology (see for example Reed and Hughes, 1992).

The emergence of *workplace studies*

Ethnomethodological enquiries into computer-supported workplace interactions have not addressed any specific debates within sociology or philosophy in relation to critiques of cognitivism. Rather, ethnomethodological enquiries emerged as a new research field, w*orkplace studies*, “in light of debates within disciplines such as Human Computer Interaction (HCI) and Artificial Intelligence (AI)” (Heath et al, 2000, p.300). Or, as Heath and Luff also put it, “in light of converging debates and developments in research on HCI,

CSCW, requirements engineering and organisational behaviour" (Heath and Luff, 2000, p.5). The topics of these debates have been, among others, the potentials and limitations of psychological assumptions, and formal descriptions of knowledge, tasks and workplace interactions.

For example, Dreyfus' phenomenologically-based critique was taken quite seriously by AI researchers in the 1990s, more specifically, the idea that intelligence necessarily requires a body to interact with the world and understand it. Consequently, AI researchers developed so-called `Heideggerian AI' (eg. Brooks and Stein, 1994; Winograd, 1995; see also Weeler, 2005), although it is arguable that the phenomenology-based critique was misconstrued in this `Heideggerian' attempt to solve the problem of computable intelligence, as Dreyfus himself has elaborated (eg. Dreyfus, 2007). There have also been shifts away from conventional cognitivism in HCI research. For example, there has been an interest in ambiguity and multiple meanings or interpretations in design (eg. Rogers et al, 2002*;* Sengers and Gaver, 2006). There has been an interest in experience as a research and design topic (eg. McCarthy and Wright, 2004; Blythe et al, 2006; see also Eggen and Kuffin, 2006; Hvannberg, 2006), and there has been a shift from studying individual cognition to studying so-called distributed cognition (eg. Hollan et al, 2000).

Studies in ethnomethodology have as well been visible in the HCI community and in organisation studies (eg. Anderson, 1994; Bowers et al, 1995; Button and Sharrock, 1995; Button and Sharrock, 1996; Button and Sharrock, 1998; Heath and Luff, 1998; Hindmarsh et al, 1998; Heath and Luff, 2000; Luff et al, 2000; Suchman, 2000b; Suchman, 2000a; Hemmings et al, 2002; Crabtree, 2004). Their main reputation is that of consistently pointing to and describing the singularness of achieving certain tasks rather than offering solutions to the problem of how to generalise what is going on. So, uncovering the particulars of some lived action and interaction has raised questions about the relevance of such findings for process engineering (see eg. Button and Dourish, 1996). One can argue however, that the findings of ethnomethodological enquiries have been particularly insightful in relation to a

fundamental challenge in the design of computer systems and supports, namely, what methods to devise for understanding interaction for the purposes of design (eg. Crabtree, 2001).

It is noteworthy that the developments briefed so far in this section have not been of much interest in the circles of sociologists or in STS. In a paper discussing the emergence of *workplace studies*, Heath et al actually remark on how little workplace ethnographies are known within sociology (Heath et al, 2000). They explain that,

> [*workplace studies*] explore the ways in which artefacts are 'made at home' in the workplace, and demonstrate how the use of even the most seemingly 'personal' computer rests upon a complex social organization; an indigenous and tacit body of practice and procedures through which tools and technologies gain their occasioned sense and relevance within workplace activities (p. 299-300).

According to Heath et al, *workplace studies* are enquiries into the uses of tools and technologies within workplace activities, as opposed to other ethnomethodological studies of work which have been centred on conversational practices and thus conversation analysis. For example, Garfinkel writes that, "[e]thnomethodological studies of work began in 1972 with Harvey Sacks's observation that the local production of social order existed as an orderliness of conversational practices upon whose existence all previous studies depended, but missed." (Garfinkel, 1986, vi). *Workplace studies*, however, emerge as enquiries into technically informed actions and activities associated with distinct complexes of equipment and, as Heath et al put it, indigenous and tacit practices and procedures. Many of these enquiries have also been conducted in collaboration with computer scientists, engineers and other specifically interested in activity supports for cooperative work. Such collaborations are evident in interdisciplinary forums centring on issues relating, precisely, to the sense and relevance of tools and technologies *in-hand* and *at-home* in the workplace. Examples are special interest groups who present and publish in affiliation with the Institute of Electrical and Electronics Engineers (IEEE) and the Association for Computing Machinery (ACM).[6]

[6] I have in mind here, for example, the office information systems group of the ACM (SIGOIS) or the computer–human interaction group (SIGCHI).

Workplace studies have also been influential in the CSCW community. We shall then look closer at some of the empirical studies, explore the ways in which issues of sense and relevance are considered, and how particular problems of working in virtual domains are addressed. But, before doing that, I will use the next section to elaborate further on questions of knowledge and skill in sociology.

Knowledge and skills

As I have already discussed to some extent, sociologists make claims about the nature of knowledge. Knowledge is a socially constituted attribute, grounded in social epistemology. But, in discussions about knowledgeable and skilled persons, these attributes are often cast in very general terms as socialised, enculturated, discursive or embodied *practices*. It is also the case that when persons cannot reproduce an accurate step-by-step description of what they do and know, or when some particular expertise exceeds that which can be made explicit by the expert, the tacit or embodied *dimensions* comprise the missing components. According to this line of reasoning, such components remain largely unseen and are inexplicable except by a generalised reference to *practice* or, as Collins prefers it, *a form of life*.

For example, early laboratory studies argue that the unique epistemological quality of *doing* expert work is thoroughly grounded in social interaction, ie., personal contact, discussion and demonstration (Knorr-Cetina, 1981; Collins, 1992). Collins takes an example of replicating an engineering know-now which, he argues, necessarily required personal contact and discussion to obtain "a crucial component of the requisite knowledge" (Collins, 1992, p.55). It could not be obtained from written resources alone.

The so-called practice turn in STS however, largely resists this reduction of knowledge to merely social interactions. Interest in technological developments and change drew attention to the role of *matter*, *objects* and *embodiment* and the practices of re-configuring environments and social-material relations (eg. Latour, 1987; Pickering, 1992; Bijker and Law, 1992; Knorr-Cetina, 1999; see also Schatzki et al, 2001). Consequently, the embodied

practices of interacting with objects, instruments and environments are an important source of explanation and theoretical grounding of knowledge—ie., as Suchman puts it, practices for which knowledgeable and skilled persons essentially serve as *stand-ins* when they are described as having expert attributes (see discussion in Suchman, 2007: ch.15). Knorr-Cetina offers us also an interesting angle on this notion of `embodied practices' when she argues that new programmers will often refuse to even try to understand the code written by someone who has left the scene (Knorr-Cetina, 1999). It is more difficult to `unblack-box' that knowledge from its traces in the code, as she puts it, than to replace it with new solutions (1999, p.98). She further argues that the body becomes a black-box of knowledge. Crucial components leave the scene when the body leaves.

To summarise, knowledge has been reduced to social interactions, explained as embedded in socio-material practices and in embodiment, or knowledgeable and skilled activities are simply referred to as *practice*. The acquisition and transfer of knowledge, explicit or tacit, is also seen to require social or social-material *practices*. There is no question that these conclusions deal a blow to individualist and cognitivist assumptions about knowledge by foregrounding what are understood to be the missing social, tacit and embodied dimensions in information processing theories, designs of cognitive control structures, and the like. But, while there is nothing particularly objectionable about a conception of tacit and embodied practices, the studies that speak of such practices typically shift the attention away from the details of lived working practices. For example, they do not attempt to enquire into and uncover the ways in which people rely on devices and each other's competencies in the course of some actual work, although, that is presumably what speaking of *practice* is meant to have a grasp of.

Ethnomethodologists use the words `tacit' and `embodied', for instance, “an indigenous and tacit body of practice and procedures” (Heath et al, 2000, p.299-300), or “the embodied practices of particular living breathing human beings” (Button, 1991a, p.5). I have yet to come across a technical use of these terms however, as of particular importance to the

business of ethnomethodological enquiry. At a closer look, ethnomethodologists challenge the convention to cast tacit embodied knowledge in terms of generalisable *in-practice* phenomena. They challenge the convention of using these terms to fill a void when practitioners give no explicit account of the circumstances of their practices—a void which then grants sociologists their claims about, say, socially governed *unconscious* choices (see discussion in Lynch, 1993, p.266). True, one does not quite clearly know what one knows and does, say, if one is asked. But people who work together rely on each other's competencies and cope with the situated singularness of respective tasks and activities, and the many ways in which competent and accountable uses of systems and supports *can* be displayed and recognised. In other words, *things* are displayed on an ongoing basis to produce and reproduce recognisable procedures and work results. The point is that many such things *can* be discovered as phenomena of competence, a skill or some knowledge—or a lack thereof.

In addition to these particular challenges to sociological conventions, it is relevant to point out that knowledgeable and skilled activities are not necessarily achieved in social settings, strictly speaking. For example, a significant finding in Jordan and Lynch's account of *doing* the plasmid prep (Jordan and Lynch, 1992) is that achieving that particular task cannot be communicated effectively by the conventional means of discussion, demonstration and printed description/explanation. It is often mastered through solitary practice and, "[i]n a sense, it is reinvented on each occasion of its use." (p.84). Jordan and Lynch observe that, since this is the case, a confusing set of issues is opened up in reference to what is commonly seen as an indispensable causal link between social interaction and the acquisition or invention of new knowledge. For example, the lone inventor was abandoned as an explanatory subject when sociologists began to study the internal social-historical contents of various artefacts and technologies. In fact, any individualist conception of knowledge has been criticised as thoroughly under-socialised and inadequate. It is clearly not entertained then if the mastering of techniques or new solutions can be privatised, although, it is arguable

that the concept of `tacit knowledge' or Knorr-Cetina's depiction of the body as a `black-box of knowledge', indicate some form of privatisation. But, where is the sociology in that?

Jordan and Lynch offer us an interesting angle on this matter by describing particular rituals and rationality in solving a set of problems whereby the technical workers have invented, in solitariness, the means to complete certain tasks. Recognising that a particular technique or solution is in *some sense* privatised however, is not the same as suggesting a turn to individualist and cognitivist conceptions. Rather, there are questions raised as to what it is in the course of solitary achievements that are matters of *social relevance*. Such questions also challenge notions of *social action* as distinguishable from action which somehow is *not* social (eg. Weber, 1947, p. 88).

On the basis of these considerations, one can say that problem solving is achieved in cooperation, coordination and in solitariness, each of which is equally interesting, ie., the particulars of such achievements if they are available. This is significant for the development of this thesis, particularly because of what is evident in the work of the operators of the archive, namely, their job required an ability to solve problems and develop operations techniques in solitariness as much as in coordination with one another. What is clearly not of interest for the purposes of this thesis however, are the theoretical problems of what can, in principle, be made explicit, what is tacit, whether knowledge can be privatised or whether a skill is strictly the causal effect of enculturated, socialised, discursive or embodied practices.

In hand and at home within the workplace

I shall now turn to empirical studies concerned with sense and relevance of tools and technologies in workplaces. Those of particular interest here include enquiries into work relating to construction sites (Suchman, 2000b; Suchman, 2000a), navigation (Hutchins, 1995b), goings-on in traffic control and news rooms, and in various industries and offices where members coordinate work on a daily basis (eg. Button and Sharrock, 1995; Blomberg et al, 1996; Button and Harper, 1996; Suchman, 1997; Heath and Luff, 1998; Trigg et al,

1999; Luff et al, 2000; Heath and Luff, 2000). The main objective of these enquiries is to uncover people's interactions with systems and supports. Empirical data gathered from ethnographies often show how activities are accomplished and made intelligible in ways that do not match requirements specifications. Real-time activities involve unforeseen contingencies and require that workers can respond accordingly in a way that, for all practical purposes, is the reasonable next thing to do on that occasion (eg. Heath and Luff, 2000).

Promoters and designers of systems tend to believe that people's practices will automatically change—they will eventually see the benefits of new systems and supports, and accommodate them (eg. discussion in Orlikowski, 1992). True, introducing new technologies can extend and reorganise work, and transform the existing work conditions (for example Kling, 1991; Bloomfield and Vurdubakis, 1994; Ruhleder, 1995; Ehrlich and Cash, 1999; Engeström, 1999; Engeström, 2000; Agre, 2000; Agre, 2003). An immediate practical problem however, is that support systems can enforce the handling of work objects, interacting with others, communicating, and so on, in ways that do not support very well the unfolding course of action. As Suchman has put it (Suchman, 1987), computational designs tend to be too restrictive and overly purposeful as if planning was the concrete foundation for an inherent rational character of our actions (eg. Winograd and Flores, 1986 and critique in Suchman, 1998). Consequently, if new systems do not fail completely, they are adapted as people figure out how to work around them and ignore some of the built-in process designs and pre-specified purposes.

In light of these considerations, it is of some interest how an information, communication and control system such as the archive can be depicted in terms of process engineering. The archive is operated to deal with continuous throughput of records and documents—ingestion, auto-processing, detection, manual corrections and dissemination. But, for any system that could by described as having those general characteristics, the obvious objectives are to develop supports (eg. objects, controls and communications tools) so that the people involved can proceed, presumably, in an orderly fashion. Similarly,

infrastructures and the various units that make up a complete system are pieced together to store, process and make accessible the work objects (eg. records catalogues, different kinds of documents and files, and so on). In other words, and I draw on my own work experience, such systems and supports are depicted as building blocks and infrastructures, but also as mediators of the activity. We shall then ask what these depictions can tell us.

The concept of `mediation' is used by activity theorists to argue that the key role of computer supports is to be mediators of human activity. In fact, ethnomethodologists, activity theorists, theorists of distributed cognition, and many others, see material artefacts - devices and what they produce - in one way or another as mediators. For example, in Hutchins' story of navigators (Hutchins, 1995b), the computational internals of the instruments produce, among other things, diagrams and chart projections which play a substantive role on the team in answering the question, where am I?. In Suchman's study of engineers (Suchman, 2000b; also Suchman, 2000a), the internals produce appearances of a construction site, of designs and intermediate arrangements, all of which crucially provide virtual access to the location where constructions are carried out step by step. The roles of these instruments are therefore to provide background operations, to be available for interception and to produce representative outputs. They support, among other things, access, tests, measures, and the construction of documents and reports.

From these understandings, one can argue that many mediating roles are *necessaril*y reliable, for example, in Heath and Luff's study of operators in a control centre (Heath and Luff, 2000: ch.4). The operators have to assume that is the case or enacting control over situations would be considerably compromised. The same is true for Hutchins' specialised teams, navigators and pilots (Hutchins, 1995b; Hutchins, 1995a). Studies of engineers and architects who use computer aided design (CAD) systems also provide discussions about already existing tool kits for well established practices (Downey, 1998; Suchman, 2000b; Heath and Luff, 2000: ch.6). There are standards for coordinates and measures. There are document standards for drawings, texts, photos and more. Standardised tools are interacted

with in order to build representations into documents and render the work visible in various formats as needed.

Arguably, the notion of mediation is somewhat helpful then. It is a simple way of stating the fact that devices and objects can reliably support tasks and tasks are delegated to them. They communicate already built-in decisions with which people interact. In other words, we can co-act and interact with *material collectives*, as O'Connell would call them O'Connell, 1993; related are also *immutable mobiles*, Latour, 1987), and with standardised facilitation for certain tasks and activities. But people also attempt to reconfigure or re-organise the devices they work with (Bowers et al, 1995; Schmidt and Simone, 1996; Twidale and Marty, 2000; Bernstein, 2000). That is to say, single objects or devices may have reliable self-contained functions, however, they can be variably situated and invoked in the course of action.

This last remark brings us to the depiction of building blocks and infrastructures which also concerns a distinction which can be made between aggregates of basic protocols and strictly purpose-built designs. Basic generic protocols agree with each other at the `lower levels', rather than strictly with the rest of the world. They serve as building blocks on which other functions can be constructed. Also, systems with high compatibility are utilised as infrastructures or platforms, such as Woolgar points out is the case with the microcomputer itself and its capacities to run applications and control periphery devices (Woolgar, 1991a). Its design is, in that respect, open ended.

Alan Dix, and other HCI researchers have made similar claims about the internet as a collection of cross-platform protocols for communication, information access, aggregation and dissemination (Dix, 1997; Grasso et al, 1997). The `HTTP` protocol came with the built-in flexibility to facilitate already existing communications protocols and it can transmit document types and formats for text, images and other media, accommodate specified information contexts, and so on. What Dix observes is that when the potential to serve as a platform or infrastructure is made widely available, practical difficulties pertaining to

restrictive purpose-built environments are bypassed in favour of the flexibility and potential for an assortment of roles. For example, he observes that the only operations running campus-wide at universities are web and internet services through which all sorts of other services come and go. Therefore it is also noteworthy that comprehensive purpose-built teamwork systems, so-called *groupware* which is studied and developed by CSCW researchers, have not yet gained significant momentum. However, sharing access to file directories and generic databases, using email, wikis and social media, are widely popular instruments in teamwork among persons of variable training in computer use, even to software and systems engineers. For example, studies of open source programmers and how they coordinate operations in virtual workspaces show us that they do not build comprehensive purpose-built media spaces to support their work (eg. Gutwin et al, 2004; Barcellini et al, 2005; Sack et al, 2006; Yamauchi et al, 2000). Rather, they rely on basic protocols that can be assembled to improvise minimally structured communications systems to exchange code, text and messages. There are discussion spaces, information spaces and implementation or operation spaces through which the programmers dynamically keep track of several different progressions concerning everyone's moves and the development of shared objects.

All of this is useful to be aware of if one is studying operations and uses of computer systems and supports—if only to have some grasp of the practical issues involved in their construction, what kinds of building blocks or infrastructures are made available and, perhaps, which ones are shown to be more feasible or more practical than others. Namely, when the practical utility and usability of systems and supports is contemplated, questions of protocol standards, compatibility, interoperability, specialised purposes, and related issues, have significant bearing on what gets deployed. These are important considerations as well in research and development aiming at better understanding `mediated' human activity—ie., the interactions and the organisation of work, as is the case in CSCW and HCI research.

Insisting on these depictions of mediation, building blocks and infrastructures, risks

being misleading however. By that I mean, for example, when building blocks and infrastructures are treated, theoretically, as stable mediators of human activity, through which stable representations of the structure, organisation and the state of the work are also mediated. One example is Hutchins' analysis of distributed cognition in navigation (Hutchins, 1995b), analysis which depends on available representations of minds and of device operations, as situated but nevertheless stable task-relevant indicators. It is not my intention to object to formal distinctions between the *general* and the *singular* when they are essential to programming efforts or to managing complex operations. The *general* is the stable, for example, the *constants*, the *variable definitions* and the many *for each* conditions in programming languages. To some extent such distinctions are also instrumental in planning for purposeful uses and meaningful practices. But they are of limited use insofar as such distinctions can substantiate how actual procedures unfold in the course of action. In other words, whereas distinctions between the *general* and the *singular* serve well to establish infrastructures and building blocks, they have severe limitations in the design of work procedures. Nevertheless, such distinctions are accommodated in research into design for workplace interactions (eg. Bernstein, 2000). But any such structural approaches ignore the ways in which system and support requirements are identified in the actual course of action as the *instabilities* of mediation and roles become apparent to the people who are doing the work. As Suchman has argued, work procedures, which often are technologically complex, train practitioners to produce and reproduce the objects and instruments of their work. They also train them to produce and reproduce each other's contributions to arrangements of mediated tasks (see, Suchman, 2007: ch.15; see also on professional vision Goodwin, 1994). There are two examples from CSCW research which can elaborate further this point. One is specific to distributed virtual workplaces and the other to issues of appropriating tools and devices.

As regards the former, CSCW and HCI researchers have been looking at notification systems or servers (eg. Ramduny et al, 1998; Gross and Prinz, 2003), video-supported

communication (eg. Ruhleder and Jordan, 1999), mixed media and part-synchronous work supports (eg. Grasso et al, 1997; Bertelsen and Bødker, 2001; Büscher et al, 2001), and various types of chat support (eg. Nardi et al, 2000; Smith et al, 2000; Muller and Gruen, 2005). These designs are developed as teamwork supports, some of which are also attempts at comprehensive support systems that should facilitate inter-connected workplace arrangements with the means – all-in-one – to communicate, work on shared objects, exchange them and make decisions about how to proceed.

One of the interesting challenges is how to support *awareness* and *orientation* within such a facility (eg. Prinz, 1999). The people involved have to be aware of each other's presence and moves, and of the general state of the work and the environment. They also need to orient toward specific distributed and shared tasks. The problem is approached as a structural problem and dealt with by identifying *lower* and *higher* level operations, *general* properties and assumptions about *singular* ones. Computable context descriptions are implemented whereby actions and the presence of and interplay with environment, work objects, instruments and others, all have distinct orienting properties. Each is encoded as if they rested on stable representations of the structure and flow of coordinated work, of divisions of labour and responsibilities. So, there are specific work contexts for the people involved, there are location contexts, contexts for responsibilities, work objects, applications associated with work contexts, and more (see also Gross and Prinz, 2003).

This and similar approaches have not been all that successful and the counterargument is that orientation and awareness properties are not necessary prerequisites to doing work. Rather, members need tools to produce a sense of awareness and of pending tasks to which they can be sensitive on an ongoing basis (Hindmarsh et al, 1998; Heath et al, 2001; Heath et al, 2002). For example, when this matter is taken up by Heath et al. (2002), they note: “that the idea or concept of awareness is predominantly spatial, like a moving beam which illuminates, in the course of action, a stable or shifting sphere of the individual's world” (Heath et al, 2002, p.319). They contest the assumption that focal points in an individual's

world are stable and discrete. Empirical data suggest partial, fragmented, asymmetrical and ceaselessly shifting awareness (also Heath and Luff, 1998; Heath and Luff, 2000; Suchman, 2000b). What the data also suggest is that individuals will produce and preserve their sensitivities for each other's presence and conduct, and selectively display aspects of their activities which are essentially localised and responding to clues embedded in immediate surroundings. In other words, according to ethnomethodological enquiries, awareness is a practical accomplishment achieved in collaboration with others (Heath et al, 2002, p.320-321). *Remaining sensitive* to a domain is then crucial rather than *being aware of it*, say, a domain for which a person is principally responsible. And, media spaces provide important resources for *becoming* aware of relevant states and changes. They *can* be used as sites of orienting and coordinating efforts while the experience shows that they do not operate very well as service channels that enforce particular orientations to pre-specified awareness contexts.

Regarding the other example, about appropriating tools and devices, it seems like an obvious demand that computer systems and supports should accommodate work and overcome, as Schmidt and Bannon put it, any support requirements of cooperative work arrangements (Schmidt and Bannon, 1992; also discussion in Berg, 1998). It is a practical demand and there are very practical and matter-of-fact considerations constantly put to the test, as Nichols and Twidale observe with respect to library practices in particular (Nichols and Twidale, 1999; also file management in Blomberg et al, 1996; Trigg et al, 1999; invoice system and crime reporting in Button and Harper, 1996). There are questions such as what tasks computational devices can realistically execute for people, which activities they can help people do more effectively, what new activities the technology can facilitate, and how we can introduce new systems and ways of working without disrupting existing successful operations.

It is noteworthy in this respect that operations internal to particular tools or devices are typically invisible to users. They never step imaginatively inside the instrumental structures

unless their profession is one in which members go about their work by reading and writing in code (Agre, 1995 cf. Suchman, 2007). Most users anticipate seamless access to the objects of their work *through* an interface—a condition referred to as transparency of the technology by which users manipulate instruments and objects without having to know how or why they can do what they do (discussion for example in Hutchins, 1995b: ch.3). Consequently, and as Suchman has argued (Suchman, 2007: ch.4), most users are faced with internally *opaque* machinery.

The general idea is then that computer systems and computer supports should be *genuinely appropriated* just like any other support tools and instruments. What these concepts of opaqueness, transparency and genuine appropriation actually draw attention to however, are questions such as *to whom* or *to what extent* systems and supports are transparent (or translucent) rather than opaque, and in what ways, if any, that has bearing on appropriation. These concepts also draw attention to the many ways in which devices are unfamiliar, how they fail to meet expectations and fail to properly support what is going on. Objects / devices also fail and need to be mended and most systems never operate seamlessly. One could ask then if this concept of *genuine appropriation* – as the *ideal* tool use – serves a similar purpose as Weber's *ideal type* of rational action, against which the uncertain, failed and fragmented uses are measured while the ways in which tools and devices are actually appropriated are more or less ignored.

To sum this up, what is at stake for my enquiry is to bear in mind, as this section has indicated, the potential complexities and uncertainties in assembling, operating and using systems and supports. If the question is raised as to how tools and technologies are made at home and in hand, in operating and using the archive, then the obvious next step is to enquire into the details of how the operators and users do just what they do in their interactions with the apparatus. As I have already emphasised, the objective is not to make generalised claims about the *nature* of information and communication systems *like* the archive or about the nature of tool use. Rather, the objective is to draw attention to the ways in which the people

who operate and use the archive make use of available supports in ways that have claims upon preference and practicality, dispersion and variation in techniques or in ways that have claims upon future tasks.

Ethnomethodology and post-cognitivist theories

I began this chapter by pointing out that research into the uses of systems and supports within workplaces often address design intervention (eg. Hughes et al, 1993; Anderson, 1994; Forsythe, 1995; Forsythe, 1999; Ngwenyama and Lyytinen, 1997; Hindmarsh et al, 1998; Trigg et al, 1999; Büscher et al, 2001; Gross and Prinz, 2003; also discussion about this involvement in Schmidt and Bannon, 1992; Suchman, 1995; Button and Harper, 1996; Halverson, 2002; Garrety and Badham, 2004). These studies commonly observe the difficulty in capturing (and overcoming at the design level) some of the basic features of group activity such as collective track-keeping, managing histories, oversight in the flow of work, relevance of devices and tools, and more. They acknowledge that work supports continue to be subject to contingencies and naturally occurring dynamics in situated interaction with them. But, as Barthelmess and Anderson argue, supports to coordinate tasks, and to an extent to cooperate and successfully meet certain objectives, have improved over time (Barthelmess and Anderson, 2002).

If this is indeed the case, then those who are involved in some capacity are learning, although, as I also explained at the beginning of this chapter, designs *as such* are not the focus of this enquiry. Rather, I proposed to take in account critiques of the assumptions on which they are typically based and the fact that studies of organisational conduct often foreground mistaken assumptions about activity, assumptions which then have been scripted into some particular design. The archive had not been purpose-built to capture group activity, collective track-keeping and what might go wrong. It was built piecemeal in reaction to what could be learnt about submitters and their submissions, and what could be done to manage preprints and bibliographic data. What is of interest then is to enquire into what counts as a

competent and accountable operation and use of the archive in spite of, or perhaps because of, the evident lack of uniformity as far as goes the preparation of preprints, available computer supports and possibilities for handling both objects and supports in keeping the documents and records in good order. So, the present task is in the main one of respecifying the problem of order, more precisely, disorder.

In pursuing various lines of enquiry, ethnomethodologists have been instrumental in uncovering how people do just what they do. But they have been indifferent, for instance, to debates on how well the alternatives to cognitivism explain what goes on—eg. the theoretical constructs referred to in CSCW as post-cognitive theories such as activity theory and the theory of distributed cognition (eg. Rogers and Ellis, 1994; Hutchins, 1995b; Kuutti, 1996; Bødker, 1996; Nardi, 1996b; Ngwenyama and Lyytinen, 1997; Engeström, 2000; Hollan et al, 2000; Bertelsen and Bødker, 2001; Ackerman and Halverson, 2004; Kaptelinin and Nardi, 2006). The critiques coming from ethnomethodologists are consistently directed at a single problem, as they see it, the `missing what'—a critique aimed at any theorising activity for trivialising `that' in the lived moment and `what' the observable-reportable is of some action, reasoning, skill, order, procedure, responsibility, etc., etc.

For example, activity theory can be summarised as a conceptual system of principles used to examine and explain dialectical aspects of activity systems which place the human subject at the centre and socially determined interactions supported with mediating artefact hierarchy. As Engeström would put it, the theory serves as a tool to reveal the instrumental conditions of routinised workplace operations, the motives behind activities and the conscious goals and needs in situated action (Engeström, 1999, p.65). Central to the theory are the principles of `mediation' and the `internalisation / externalisation' (also unity) principle of minds and mediated conduct. Also, the theory of distributed cognition can be summarised as an analytic framework designed to explain action and the organisational order on the basis of representational media and their propagating states. The complete unit at work is a cognitive computational phenomenon, eg. “the ship navigation as it is performed by a

team in the bridge of a ship" (Hutchins, 1995b, p.49). The theory serves as a tool to examine and explain the observable processes, manifesting what goes on inside both minds and devices, more precisely, the observable processes between them.

Both these theories can be read as powerful critiques of information processing psychology in HCI research for emphasising social, cultural and historical contexts of work. They also acknowledge embodied situated interactions between people and with material artefacts. Hutchins makes this explicit while, nonetheless, he also states that he uses normative accounts to provide a stable framework for describing the structural properties of the social organisation (Hutchins, 1995b, p. 178; see also discussion in Heath and Luff, 2000, p.88-89). Activity theorists also insist that regardless of interactions being embodied and situated, it is essential to examine and explain what drives people and not lose sight of motives and goals (Engeström, 1999, p.64). Both theories are actually grounded in psychology. Activity theorists speak of socially determined interaction which draws on cultural-historical psychology and a development theory of cognition (Vygotsky, 1978). And, as Button has pointed out, Hutchins makes his claims using the terminology of cognitive science (Button, 1997). The social organisation is a computational architecture and culture is a process both inside and outside of minds (see also discussion in Heath and Luff, 2000, p.16-17).

Finally, both these theories, as well as other alternatives to conventional cognitivist assumptions, have a curious problem insofar as they are developed to impact design decisions. They are at best theoretical approximations (see Kuutti, 1996 on activity theory in particular). As Halverson puts it, they lack prediction and prescription. "What a theory can warrant is not all that is necessary to make it useful" (Halverson, 2002, p.261; also response in Nardi, 2002). This is an unfortunate problem, given the stated intentions, but it is not an unfamiliar one. Social theories lack in prediction and prescription on how social structures are actually constructed. But, as far as I can see, it is not on the question of prediction and prescription that studies in ethnomethodology break away from sociological or psychological

theorising and, indeed, avoid theorising in a studied fashion. The matter is perhaps to some extent a response to less-than-credible sociological and psychological explanations, but mainly it is one of radically different study interests as I discussed in the previous chapter.

An interesting comparison is also made between CSCW research and workplace studies in an article that articulates how ethnomethodology and activity theory differ (Suchman, 2000b). “This difference turns on their relations to the project of theorising itself” (p.15). Suchman quotes Nardi's argument in the article—which I quote in part here—that “[a]s we move toward ethnographic and participatory design methods to discover and describe real everyday activity, we run into the problem that has bedeviled anthropology for so long: every account is an ad hoc description cast in situationally specific terms.” (Nardi, 1996a, p.10). Suchman questions whether this issue of the “ad hoc, specifically situated character of accounts is a special problem for the social sciences that renders comparison impossible in the absence of a unifying theory” (Suchman, 2000b, p.15). To quote Suchman further on this note, “[t]heory building and theories-in-use are themselves specifically situated activities” (p.17). As ethnomethodologists would argue, the problem Nardi points out is not one for which sociologists should try to find a remedy.

I conclude on a similar note I did in the previous chapter insofar as this chapter has continued the work I set out with—to clarify my relation to the project of theorising and explain the study interests introduced by ethnomethodologists. In short, ethnomethodological enquiries do not engage in constructive analysis of what governs actions and organisational life or how objects and devices are appropriated and used in organisational interaction. Rather, the distinctive craft is to describe site-specific phenomena of action and interaction intuitively recognised as practices *in-their-course* in settings which are understood to be self-organising. I will now move on to the next chapter where I will foreground the central concerns of this enquiry. I will describe the available data and what I have done with it, and relate my choices to further discussions of social science methods and the study policies of ethnomethodology.

Chapter 4

Central considerations

Introduction

In the previous chapter, I argued against the idea that electronic building blocks and infrastructures can assimilate the `building blocks' and `infrastructures' of computer-supported work. The former are conceptions of *general purpose* and *specificity* which contribute substantively to configurations of systems and supports. But the latter, if thus conceived, cannot adequately contribute to configurations of meaningful and useful work supports. I also argued against the idea that so-called `background assumptions' characterise mental states *prior* to action, or that motives, goals or skills can exist *independently* of the actions and interactions *in* and *through* which motives, goals or are recognised and realised. The studies that are based on such assumptions are of limited use in solving the practical problems of appropriating work supports or understanding *systems-in-use*. Furthermore, Chapter 2 argued against the idea that social structures, forces, histories and circumstance stand *behind* or *surround* events and the actions of individuals, and thereby serve to explain order as *caused* in that specific way. I argued that I have yet to see the evidence of such `functions' in everyday practical action and reasoning (see eg. Garfinkel, 1967: ch.3).

It is not as if ethnomethodologists are wishing away structure or the social order for analytic or other purposes. They are not arguing that actions and events can happen *completely* independently of such phenomena. Nor are they arguing that methods of classifying and ordering work in a structured fashion are useless, whether or not the work is computer-supported. Rather, from an actor's point of view, some form or another of order, knowledge, skills, etc., are always already relevant to the purposes at hand and that *world* which is always already there can be consulted if events and actions are called into question.

The thrust of this argument, in my opinion, lies not so much in the ways in which a long history of ontological and epistemological assumptions, manifested in cognitivist or social-scientific explanations, is subject to doubt. Rather, as I emphasise in the previous two chapters, its thrust lies in the attention drawn to different study interests—to enquire into lived practical reasoning *in* action and interaction. The natural assumption to go with then is that goings-on will not and cannot proceed in a meaningful way if the respective settings are utterly unintelligible.

What I will do in this chapter is to narrow considerably the focus of my review and clarify methodological issues which are immediately relevant to this enquiry, what data were available, how they were recorded, and what policies I will use for selecting cases. I will explain how it is that formulating pre-given methodological procedures for selecting and analysing data is not the basis on which the goings-on at the archive will be described and explicated. Rather, there are topics of (dis)order which I wish to treat procedurally and thematically, those of *inspection*, *detection*, *corrective action*, and the use of supports. I will come back to this matter.

For the present purposes, I will first continue the story I began telling in the opening chapter. I explained that `bitnet`, `TeX` and preprint exchange were considered essential to the founding of the archive and I elaborated how these `ingredients' came about. I will now describe some of the technical arrangements for operating and using this system. In doing so, I rely on my practical knowledge and experience as anyone would posses in the capacity of an operator on site. I also rely on what the other operators had to offer during my time on board and what the inventor has put on record over the years.

When researchers do not have the practical skills of those whose situated practice they study, the potential shortcoming of their enquiries is to have nothing to say about the contents of that practice (eg. Lynch, 1985). But to do the opposite and engage at the level of other participants demands resources. In that respect, there is no question that my participation is a given in this story. I believe that I possess the ability to adequately describe the technical

arrangements. I will also attempt to describe and exhibit recognisable features of procedures even if the reader is only minimally acquainted with system operations among so-called *for-free* and *open-source* enthusiasts, often referred to as `hackers'. I will rely on what Lynch calls *vernacular expressions* (eg. Lynch, 1993) and *common sense* which can relate to co-membership of peers and other readers as a good enough basis for my claims to understand and communicate this material (see also discussion on Lynch and Bogen, 1996 in Ten Have, 2004, p.37).

This chapter will further elaborate on issues of *common sense*, *everyday* and *technical* expressions, but the next two sections will draw attention to the shared work space on the archive and why it was there in the first place. I am referring to the `place' where the operators would keep track of each other's involvement in everyday goings-on, the communication with submitters of preprints, and where they exchanged their comments and ideas on a range of events and issues relating to the operation and use of the archive.

A mass-dissemination system

As I explained in the opening chapter, the archive has served the practical purpose of sharing (electronically) the latest preprints in physics and mathematics. It started as a simple email reflector that could only be accessed from remote on a command line (Warner, 2001). `TeX` files and bibliographic data were uploaded by their authors via the `mail` protocol onto a server machine and from there they were instantly available to anyone to download and compile on their own `Unix` computers with `TeX` processing software.

In the first two or so years, the founder of the archive single-handedly ran the operation on his `Unix` office computer and often left the system to its own devices. There was no quality control of the bibliographic data, no auto-compiler of `TeX`, and no one in service to submitters and readers of preprints. Soon it was known, however, that many of the `TeX` files were not compiling without problems and the founder's opinion of it was that it was up to the

authors to fix these problems. After all, the system was meant to be fully automated. It was set up to be a *pure dissemination system* (Ginsparg, 2003)—ie., a system that automatically ingests and disseminates digital objects.

Automated data processing typically refers to some form of mass-processing where, ideally, any `broken' objects are insignificant outliers that can be deleted or ignored. But this means that there is always a question of the error level one is happy with. For example, in the business of ordering documents and records for digital library services, there is really no level of error acceptable. Each bibliographic record should always point to the correct object, and both records and objects should be accessible error-free. That would be the ordinary business of managing library holdings and this is what electronic systems are set up to manage with a few persons in the loop whose job is to upload data objects, quality control records, and so on. But things were of a different order on the archive, the most obvious difference being that the archive is a self-archiving tool for thousands of authors of scientific articles. From the very beginning the authors were expected to take responsibility for their preprints, to upload error-free files and supplement the preprints with adequate bibliographic data.

One of the first things I noticed when I came on board was that the archive was not a purpose-built digital library management system. The system that I came to know had been put together piecemeal and *ad hoc* from the time it was a simple email reflector. It was in many ways incomplete and it was meant to be incomplete in the sense that the operators worked the archive as a laboratory for new information and communications technologies, for information retrieval research and more. This work was done by way of *for-free* and *open-source* software development. Commercial viability was not an issue. The archive was operated in an academic research environment. Rather, the viability of particular technical protocols and systems engineering were considered priorities over any hands-on support to submitters. A fragment from an auto-response generated upon successful author registration with the archive spells out what the priorities are:

```
Staff time here is dedicated to improving the software and adding features,
and tuning the server and mirror network, rather than assisting individual
users with minor problems that can be solved entirely at the user end.
```

The operators had set up a website and a web submission module after the `HTTP` protocol became available. A `TeX` auto-compiler had also been pieced together so that `TeX` submissions could be processed `on the fly' and `postscript` outputs (later also `PDF`) made available directly for download. But when the auto-compiler was installed and tested in the mid-90s, quality control of incoming `TeX` files became necessary in order to identify problems that could either be associated with the workings of the auto-compiler or otherwise sent back to the authors for correction. Old `broken' submission packages also came under scrutiny. It was not so much that the older preprints needed to be available for ongoing scientific research. Rather, the idea was that they should be made available since this rapidly growing digital library was acquiring historical value. In other words, outliers could not be ignored completely or simply deleted.

New features and adjustments had been implemented in response to goings-on over time. As the use of the system grew, an increasing number of `TeX` versions became available and new versions of templates for formatting the scientific texts often lacked in backward compatibility. Document and records handling technologies, figure manipulation and rendering tools also developed and changed over time. What should not come as a surprise then are the variations in preparing preprints in `TeX`. There are variations in how to post-process data plots into `reasonable' figures or images, if and how to insert figures into the scientific texts, and how authors can format author lists, institutional affiliations, and other bibliographic data. It also became possible to prepare `TeX` files with `Microsoft Scientific Workplace` and, for some types of articles, to simply use `Microsoft Word` and then export to the `PDF` format. Consequently, fragmentation in the daily operation of the archive should not be much of a surprise either. There was more and more to look out for, ie., a *growing variety* of problems more than a *growing number* of them. Data processing on the archive needed

systematic inspection, interception and repair in coping with the outliers but the *ad hoc* manner in which the instruments to do so were implemented meant that there was no single comprehensive administrative environment, nor were there any hard and fast rules in dealing with authors and their preprints. Generally, the operators were explicit about their efforts to minimise service labour on their end and leave the work to the submitters, for example, by removing submissions if they were not up to standard, pointing out what the problem was, where to find instructions so that authors could resubmit, or simply by ignoring the authors for as long as possible, anticipating that they would sort out the problems by themselves. In other words, there was no clear cut way of coping with the fact that this mass-dissemination system was not as *pure* as the founder had envisioned it.

Working together inside a message box

As I explained in the opening chapter, I will not be treating variation, fragmentation, conflict or ongoing innovations as sources of disorder but, rather, as pre-theoretical phenomena that should be enquired into. Consequently, if the daily processing of preprints on the archive required inspection, interception, testing and repair, it will be imperative to focus on the actual site where the people involved in doing this work can access and communicate what is going on. But the idea of an `actual site' of archive operations is somewhat confused.

> ...[S]omething that I had barely anticipated before I started the research, was that they [programmers] would sit for hour after hour in front of their terminals barely uttering a word. To ask them a question would be seemingly to break their concentration with the machine, to disturb the peace of the office. Even when sitting inches away from them, I was to learn nothing about the implementation. [...] Much later, however, I would realize that even while we had sat there in silence, they were in fact speaking, sometimes shouting. Their method of communication was electronic mail. It was this realization that they were in fact talking in the main via e-mail (sometimes even preferring to e-mail the person sitting across from them!) that led me to begin to sift through old, archived messages. (Pollock, 2005, p.502).

When Pollock realises that he should be sifting through archived email messages, it has dawned on him that, whatever is going on in the work of these programmers, his access to

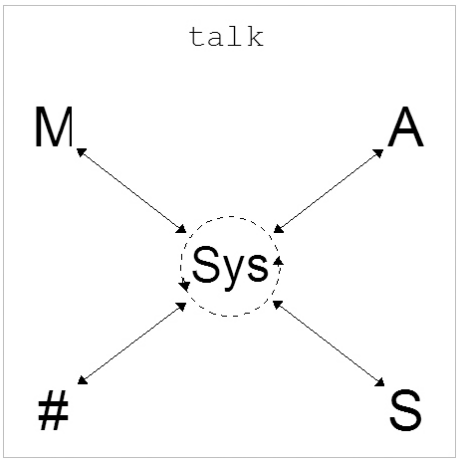

Figure 4.1*: Textual interaction and communication*

goings-on is mostly confined to logs of electronic communication. As he states in the quote above, “they were talking in the main via e-mail”.

The ways in which work was communicated on the archive was also via email. The daily administration of new preprints did not have a physical location in the conventional sense. The operators did not occupy the same room or even the same office complex. Actually, each of the operators was working in physical isolation, sometimes a few time zones apart. Consequently, they did not talk to each other while work was under way except via email. Communication with the users of the system was also via email. Messages from them were routed into a `message box' and all the replies were automatically copied into it as well. Thus it was always readily available to the operators what was communicated to submitters, moderators or others who contacted the administration.

The electronic exchange was also used for other things than to substitute natural talk in the course of sorting out problems and getting the daily work done. It was used to transmit and make visible events and actions, and to share any immediately task-relevant information. As figure 4.1 illustrates (abstractly), routes for data transmission were set up to service all the necessary interaction and communication. There was an internal exchange between a range of system server (**Sys**) components and data processing units, for example, for processing TeX files and records with author lists and affiliations, abstracts, and more. There was an internal-external exchange as well *through* the system server or *with* one or another of its components. The data processing units would report on progress and problems to the *administration* (**A**), and to relevant *submitters* (**S**). The operators would inspect and keep track of the actions of submitters, the involvement of *moderators* (**M**), communication with

miscellaneous *others* (#), the daily throughput of preprints and records, and of each other's care-taking of objects and services more generally. If they found something amiss they would investigate and report as necessary.

It should be emphasised that imperative in all this is meticulous reporting of events and actions with the support of various system components. For example, server machines have always had built into them some logging of their performance and of user/administration activities. It is part of standard practice. Scripts are written to detect and report known potential threats simply because no human would ever be able to manually look over all the interactions on a busy server machine. Thus, the built-in track-keeping and reporting on automated procedures on the archive supported necessary *observation* and *description* by the system itself for the operators to consult and act on if necessary (see Knorr-Cetina, 1999, p.55-62 on a similar matter). Moreover, once human-to-human communication is taken online, it becomes a form of data transmission which lends itself to the sequential tracking of reports and communication.

In the absence of an actual physical workplace, this arrangement was the only means for an operator on the archive to adequately keep track of what was going on. It was not sufficient enough that each operator had direct access to the insides of the system server (devices, objects, controls and their source code). It was not sufficient either that each of the operators could privately give support to submitters, moderators or others who were interfacing with the `outer layers' of the system (interacting with submission modules, communicating with an operator, accessing information and documents). If no one knew who was attending to what and how, the daily administration would have been incomprehensible and impossible to sustain. From their local terminals, operators could and would inspect and intercept data objects and devices, implement new designs or configurations, modify software and basically hack the system as they needed far beyond the daily work of administering new preprints and being of service to users. But the only *collective space* where the operators on the archive actually `met' each other in doing the work that needed

doing on the day – the space where events, actions and communication were made visible – was a message box. For that reason, particular goings-on that I wish to enquire into happen *inside* the box.

Participating, describing, explaining

Before I discuss the available data and a policy for selecting cases, I will use this section to briefly describe my own participation on board the archive and further state my claim to the aspiration of *unique adequacy*—ie., in order to understand and describe a practice it is desirable that the researcher is able to perform it competently (Garfinkel, 2002). I quote Lynch on the matter when he says that,

> Ganfinkel's "unique adequacy requirement of methods" should not be understood as an admonition to learn the discipline investigated as a prerequisite for an ethnographic analysis. Although it does warn one away from the privileges of grand theory and speculative criticism, as I understand it, the requirement has to do with a method for demonstrating what a description says about a practice by enabling readers to see what is said by entering into the phenomenal field of that practice. (Lynch, 1993, p.302).

So, to follow Lynch, the present project aims at "enabling readers to see what is said" *and done* on the archive.

Most of my own time on board was spent in the daily administration of new preprints. In doing that work, the message box I discussed in the previous section was always in my view along with three other terminal windows on my screen, all through which I carried out my share of inspection, detection, corrective action, and using supports. I did this for close to two years and thus acquired substantive know-how of what to look for in the daily grind, as the operators called it—how to look and discover what to do next.

A couple of observations by online ethnographers are familiar. Hine points out how the presence of other humans is genuinely felt inside virtual environments (Hine, 2000). One anticipates the presence of others, that someone is reading one's message, looking at the same report, the same communication. A response or some contribution from others is always imminent. It has also been noted that in the absence of physical context there are disruptions

in what otherwise is normal turn-taking. Simultaneous feedback is missing when perhaps it is strongly preferred, for instance, under time-constraints or critical conditions (eg. Herring, 1999). I subscribe to both of these observations. As to other common concerns in comparing virtual field sites and conventional research sites, I have reservations. The differences between the virtual and the real are explored and assessments typically made with respect to what can be observed or recorded, how the boundaries are drawn of the site of observation and the social setting. Questions are raised if and how virtual communication is representative of a community, a practice or a participation in personal and public affairs, compared to the ways in which such matters have previously been known to social scientists (eg. Markham, 1998; Jones, 1999; Wittel, 2000; Williams, 2006).

None of these considerations are of interest for the present purposes. The purpose of this enquiry is not to privilege theory or speculative criticism on community or organisational life on the basis of the work that needed doing on the archive. There are already competing attempts in that vein, for example, attempts to frame an organisational structure into which hacker communities can fit or to use metaphors – *imaginizations* according to Morgan (1986) – as a way to imagine and characterise organisational interactions among hackers (eg. Raymond, 1998b; Raymond, 1998a; Barthelmess and Anderson, 2002; Gutwin et al, 2004; Ducheneaut, 2005; Sack et al, 2006; Görling, 2007).

For example, concepts from actor network theory have been used tie together the negotiations, cooperation and control, aggregation and dissemination of objects, into *networks of associations* (eg. Tuomi, 2001; Østerlie, 2004; Sack et al, 2006). Explaining hacker practices in the context of larger social bodies has been attempted with reference to symbolic interaction and social formation (Lehmann, 2004) and the Lockean theory of land tenure (Raymond, 1998b). Among other characterisations are meritocracy (Raymond, 1998b; Lehmann, 2004), heterarchy (Iannacci and Mitleton-Kelly, 2005) apprenticeship (Sack et al, 2006), and sportsmanship (Görling, 2007). The last on the list, *sportsmanship*, describes the enthusiasm of for-free and open-source enthusiasts who are working for free, giving their

inventions away and climbing the ladder of reputation and influence by demonstrating their capabilities as hackers (Görling, 2007; also Nuvolari, 2005). They are frequently observed openly sharing what they are capable of which researchers describe as ego satisfaction (Raymond, 1998a), chiefdom (Lehmann, 2004), self reliance and self-management (Coffin, 2006; Crowston et al, 2007). Along these lines, Sack et al come to the conclusion that "[s]uccessfully demonstrated technical proficiency is the primary key for status ascendancy - but not the sole indicator" (Sack et al, 2006, p.237; also Raymond, 1998a). Status ascendancy, they argue, takes on a form of apprenticeship and the structure of this process resembles that of guilds. But if the focus is directed at the technical labour, it is typically treated - as Sack et al do - as cognitive activity on the basis of cognitive ergonomics, distribution and the like.

I have no objection to the use of analogies or metaphors if they can contribute meaningfully to a description of what takes place. In that vein, I can describe the operators and users of the archive as `shepherds'. They are keepers, guardians, protectors and defenders of documents, data and code—ie., the `sheep' which need to be kept in order. But this is not so important. It is the tradition of avoiding what *can* be discovered in the lived details of the technical work that I wish to break with. Therefore, I will focus on *enabling readers to see what is said and done*, although, as Lynch points out in his treatment of the unique adequacy requirement, only a few studies in the ethnomethodological corpus ever achieve that by way of textual description and demonstration (Lynch, 1993, p.302).

Available data

None of the data that were readily available are researcher-provoked as Silverman puts it (Silverman, 2006: ch.6). There were no systematic attempts to elicit questionnaire responses or devise semi-structured or non-structured interviews. Rather, event logs and reporting, often including or referring to pre-written messages, support documents and code, were collected as a matter of course. These data are textual documents of one or another kind and,

for the present purposes, I distinguish four types of text for the sake of clarity, ie., in order to explain in this section the different function each `type' had in the daily administration of preprints.

(1) Shared personal/official electronic interaction and exchange (the log data)
(2) Pre-written messages for official/public communication
(3) Global/local support documents on the website and the intranet
(4) Code (`Perl`, `TeX`, `Shell` commands, etc.)

The data are all digitally born and the vast majority are the log data (1). By default the log data accumulated chronologically on my `Linux` desktop machine. I have already discussed the circumstances of their production. The logs were produced by system components and the people involved in the operation and use of the archive. They were produced in the course of *doing and communicating work* or any issue involving a new preprint. I consider these data a *recording* of what took place inside the message box. So, rather than speculating on what the logs actually represent, one can argue that the message box constituted a hub through which events, actions and interactions passed and were mechanically recorded.

References and quotes from pre-written messages and support documents in the log data, constitute types 2 and 3 on the list above. Paragraphs were pre-written to save time and they were used to inform submitters about some known aspect of using the system, about things that can go wrong, and more. They typically contained advice, often with a link to a help page on the archive website:

```
We do not accept low quality PS using bitmap fonts.

Your submission looks as if it was produced from tex. If
that is the case then you must submit the tex source.

See http://archive.domain/help/faq/whytex.html
```

These messages were regularly visible in the message box because the operators used them as *canned* responses. Some of them were also routed by the system into email auto-replies that went `silently' out to submitters who then quoted them in communicating a problem to the daily administration.

The support documents were either locally or globally accessible from the archive web server. The local materials were access-restricted on the intranet and available only to the operators. They were reminders and hints about set-up and configurations for the server machines, local terminal machines, mirror network, and many of the established procedures for operating the daily administration of preprints. Globally accessible support documents were available on the archive website and written for submitters as technical instructions. They explained uploading procedures and provided a list of frequently asked questions (FAQs) and the answers to them. The FAQs were often referred to or quoted in the daily administration as in the last line of the example above.

Finally, operators and submitters frequently quoted or referred to some code in their communication (4). The reason for this is that a reference to some visible process, a command or a function in a script, might be essential in the handing of objects, devices and controls, eg.,

```
If you look at pscheck you'll see that it does:
foreach f (*.ps *.eps *.PS *.EPS *.epsi *.EPSI)
```

In other words, discussing and working the code was integral to the work of preparing, submitting and administering preprints [Appendix II contains information on preparing data and the convention for the use of notations and representing different types].

Personal and professional accounts

This study was possible because of my know-how of archive operations. I worked as an operator but I also worked on analysis and planning for alterations in the way the daily administration of preprints was organised. I kept track of all my email communication and I kept notes of events. However, it would not be accurate to label the methodological premise for this project *ethnography* in any conventional sense (Hammersley and Atkinson, 1994) nor is it strictly *participant comprehension* (Collins, 1992). Rather, it is on the basis of my

personal and professional association with the archive that I generated accounts that could be later treated as documents.

On the issue of participation and comprehension, I was already involved as technical personnel before I had the opportunity to think about the archive as a case study. I was well-versed in the kind of engineering environment as the one used to run the archive, and I knew from first hand experience what tensions potentially arise in relation to design, management and maintenance of complex computing systems for academic resource services. Although tensions in the workplace are made visible in this thesis, my participation was not aimed at analysing a controversy in the information science and technology domain. There are no competing claims in this story of what is the case or how technical matters such as design and implementation are best solved. There is hardly anything said either about the end to which the apparatus is a means as a preprint service or about any controversy that might centre on the issue of openly circulating preprints.

In retrospect, the transition from being on the job and working on a thesis was sketchy. While I was still on the job nothing much happened as regards a systematic research question or a particular research hypothesis. All my probes had immediate practical or procedural relevance to doing a job. Once I left, I continued to entertain the idea that I would find patterns in the data that corresponded with a reasonable and interesting hypothesis, a theory or a cognitive map. This was a phase through which I felt as if I was grappling with chaos and that my job was to install order. My only answer to the problem was to keep reading the literature and come back to the data over and over again, usually probing them according to my latest reading. I knew though that if, in fact, the data illustrate chaos (in particular the log data) there never would have been a successful processing and public posting of over 36 thousand new scientific articles in the year 2002 alone. I knew the data were not meaningless or somehow not orderly.

From my notes, I hand-picked anecdotes, a few remarks from co-workers and a scene from one meeting with all members present who had a stake in archive operations at the time.

These data are selected to help set the stage for the occasions I am exploring, to inform the reader and offer reflections on incidents shared with the team. As to other choices I made, I quote pre-written messages to submitters and available documents that are quoted in the exchange with them. These are selected on the basis of showing what is communicated about the system, its functions and its limits. I have replicated standard basic preparation of a preprint document file, typeset in TeX. I also quote dynamic system auto-reporting to show how they contribute to the ebb and flow of work and how they communicate actual and potential problems.

As I have stated, the work that needed doing on the archive centred on handling and coping with phenomena of disorder. It was after all the presence of outliers that originally brought my attention to how the labour was constituted. I was looking at how the operators monitored data traffic and how they used home-made command-line devices to handle things that were not in order. In my early days as an operator, what stood out was all the `tinkering', the `intimacy' in communications with the archive apparatus and the discord with existing practices at the library. But the work in the daily grind was executed in isolation and thus incredibly difficult to describe to my library colleagues. To overcome that problem, we agreed that I would produce a self-recording of a single day while working as an operator to have a better understanding of how ideas for redesign could be discussed and planned.

Although this voice-recording was not made for this thesis, I made the choice of using extracts from it because it gives a rare insight into the solitary work of the operator at the interface [see transcript in Appendix IV]. In this recording, I describe what I am doing, what I am seeing and I make spur-of-the moment comments about the activity as it unfolds. This was taped on a dictaphone and later transcribed [see description of the conventions and notations used in the transcript in Appendix II].

Documents as data

If the argument is made that these data are after all factual documents or records then, as Ten Have explains, they will be considered some form of evidence of events. They are either the means to access original events or they are specimens of their own type, whereby the practices of documentation can be studied as ways of using documents (Ten Have, 2004). We assume then that documents are *social facts* in the sense that they are deliberately produced, distributed and shared in organised ways for one or another reason (eg. Silverman, 2006, p.168; also Scott, 1990). They are produced and reproduced to assist in some activity and the investigator should thus be sensitive to the conditions of producing them, how they can `fix' certain aspects of current events and actions, and what the conditions are for making them available, ie., how information *travels* through time and space. It seems therefore that the two are intimately tied, documents and the practice of documentation. According to Ten Have, any serious study of documents leads to a study of documentation (Ten Have, 2004, p.104).

There are numerous questions raised about documents as factual or as specimen. Are they credible? How well or sincerely do they represent events? Are they representative in the sense that they serve as representative samples of a totality of documents? Questions are raised about the practice of documentation, about bad, incomplete or biased records. Practices of documentation are typically in service to local practical purposes and the usability of such records for research purposes can therefore be at odds with researchers' plans to make use of them in a factist manner to study that organisation (Ten Have, 2004, p.96). In any event, documents always represent something, some version of reality, by referring to events, objects, persons or ideas—representations which then are available for consideration in the same or some other context. As Silverman points out, this understanding of the representative capacity of all *text* underpins standard approaches to analyses of content, rhetoric, narratives, themes, discourse and conversation (Silverman, 2006: ch.5; see also eg. Fairclough, 2003).

In light of these commonly raised considerations about documents as data, I argue that

the data I described in the previous section need no particular justification as documents insofar as they are recognised as relevant to enabling readers to see what is said and done. There is no reason to cast doubt on these documents as *sound* in terms of originality, *authenticity* as in who actually created them, *sincerity*, *accuracy*, and so on. They are an instantaneous production of practical devices for immediate reference and use. They bypass conventional factist or specimen considerations once their construction and use is seen as being integral to routine activities as shown in *workplace studies* (eg. Blomberg et al, 1996; Button and Harper, 1996; Trigg et al, 1999; Luff et al, 2000). Any problem of soundness, authenticity, sincerity, accuracy, rhetoric or meaning would be effectively resolved as a matter of course in the very construction and use of the selfsame documents. I will now turn again to the issue of representativeness and look at its manifestation in ideas of measurement, inference and empirical generality.

A scientific method

The early modern sociologists advocated a natural science model of sociology. As I discussed in Chapter 2, they inherit philosophical characterisations of empirical science with the uncritical assumption, as Benson and Hughes point out, "that the superiority of science over other forms of knowledge was its use of *a* method, in a word, the scientific method." (Benson and Hughes, 1991, p.111). Benson and Hughes continue to explain how scientific method is seen to connect empirical evidence to the truth status of theories. It establishes a connection between particular empirical events and general theoretical claims, and its success (and that of the natural sciences more broadly) lies in the mathematisation of nature.

The previous two chapters discuss critiques of sociological and psychological reasoning, but without addressing the issues that are specific to collecting and using figures, setting up measurements, developing coding schemes and conceptual frameworks. The previous chapters do not discuss how quantitative and qualitative studies are polarised into opposite camps or how statistical and case studies have been pitched against each other on the issue of

the generalisability of findings (eg. Stake, 2000).

To summarise, methodological discussions in the human and social sciences either attack or justify particular stances. But two problems, for both quantitative and qualitative approaches, are systematically avoided. One is the problem of securing *adequate* empirical reference, as ethnomethodologists have argued (eg. Button, 1991b)—ie., social-scientific methods, quantitative and qualitative, case studies, comparative studies, and more, make it difficult, if not impossible, to perceive or reconstruct the phenomena as they would be encountered in real life, prior to the adoption of theoretical and methodological prejudices. The other problem is that of philosophical scepticism. Sociologists avoid the fact that it is only in the absence of systematic doubt that people can proceed, for all practical purposes, with what are irreducibly events-in-a-social-order. Social-scientific methods demand that we systematically doubt and ask whether there is not a *stricter* and *more demanding* level of enquiry to get at a *deeper* understanding, although learned reasoning about the truth statuses of social phenomena is inseparable from everyday intuitively observed truths.

Theoretical endeavours have been mathematised in the natural sciences and, consequently, connections between the particular and general are secured mathematically. Attempts to do the same with human or social phenomena, processes and patterns, are not unproblematic. Such attempts are faced with difficult problems of how to reconcile relationships between everyday and technical concepts, and how to secure empirical reference. We shall then look further at these attempts at calculability, empirical specificity and generality.

An apparatus of descriptive accounting already existed prior to the early social sciences, ie., the use of figures which later became known as official or moral statistics on mortality, imprisonment, unemployment, unmarried persons, and more. As Benson and Hughes describe it, figures and their covariation were taken,

> as the evidence from which inferences can be drawn about the real social processes and patterns which the figures, though not in an unmediated fashion, reflect. It is this move from what we might term a descriptive to an analytical use of quantification

> which presages the orthodox machinery of social research (Benson and Hughes, 1991, p.112-113).

Most of the now familiar techniques are borrowed from a range of other disciplines. But thinking carefully through questions of how human/social action is empirically investigated has been paramount in securing a sound basis for dealing with evidence and inference in sociology.

Benson and Hughes take the example of traditions based on the work of Lazarsfeld (eg. Lazarsfeld and Rosenberg, 1955; Lazarsfeld, 1958). Discoveries of patterns in social research data need to overcome the fact that social phenomena are qualitative in character which means that social actors and other units of analyses are treated as `objects' with specifiable properties or attributes. Object descriptions tend to be vague, however, such as the topics of *class*, *expert*, *labourer*, *consumer*, *user*, or qualifiers and quantifiers such as *good*, *excellent*, *sometimes* or *often.* Object descriptions are nonetheless represented as indicators although such indicators could in most cases stand in for a collection of variable attributes. Also, the measurements are always indirect rather than arising from the structure and character of the phenomena itself as in, say, mathematics. Consequently, researchers establish relationships between indices which at best represent probabilistic relationships in the actual lived reality the selfsame indices represent. The stability of the outcome depends on criteria derived from the methodological techniques themselves and any patterns thus discovered are not laws in the sense that laws are introduced by natural scientists. Rather, they are empirical generalisations and law-like statements associated with a sample or a particular group under study. As Benson and Hughes put it, we learn very little about the underlying phenomena and their properties, only what they look like through the imposition of a format (Benson and Hughes, 1991, p.121).

Although mathematised measurements are not an objective in qualitative research, quantitative and qualitative methods have a number of things in common. Qualitative studies also search for patterns to facilitate explanation. Topics such as *class*, *expert*, *labourer or*

consumer are given unified treatments. They are subsumed under overarching conceptual frameworks using analytic categories such as *values*, *norms*, *order* and *interests*, defined to apply over time at the level of large social bodies (see eg. discussion in Lynch, 1993). Both search for principles that can sort out the reasons for conflicts that challenge the *common sense* of everyone involved. Both establish an ideal *standpoint* from where judgements about the research setting and the data can be made. For example, in ethnographic studies the researcher is expected to sustain his or her *strangeness* rather than *going native*. This is seen as essential to a so-called *stepping-out* process for the writing of representative text (eg. Hammersley and Atkinson, 1994). But ethnographers, working in the interpretive tradition, recognise how difficult it is to draw such boundaries. They are unavoidably at home in some ways in the environments they study (discussion eg. in Wittel, 2000). As Coffey points out, the distinction made between strangeness, on the one hand, and over-familiarity as its opposite, is too simplistic (Coffey, 1999, p.158). Although ethnographers emphasise the difference between being a spokesperson for a group versus an examiner of it, precisely to have a *standpoint*, the examiner's *point of view*, they also acknowledge how prolonged presence on site will always underscore both strangeness and familiarity (eg. Coffey, 1999). Finally, both quantitative and qualitative approaches are descriptive apparatuses which can reproduce recognisable features of the lived world, for instance, some of the *thick descriptions* constructed by ethnographers (eg. Becker, 1963; Traweek, 1988; Hutchins, 1995b; Williams, 2006). In doing that, they also share with quantitative studies the problem of establishing sound methods of *interpretation* and how to reconcile common sense everyday concepts and the technical concepts of their respective analytic formats.

One question raised then is how recognisability is descriptively achieved which also raises another question about this unavoidable use of everyday concepts to describe social phenomena, some of which take on an elevated role of theoretical specificity. The latter question regards *common sense* versus a *sociological sense* concerning topics and resources of enquiry and the methods of making inferences about general characteristics from particular

ones and the other way around. The former question, however, concerns ethnomethodological studies of sociological methods and reasoning as practical achievements. While measurement, inference, evidence, theory, and the like, are the building blocks of the prejudices of sociological investigations, ethnomethodologists have taken interest in how these matters can be respecified in the details of their workings, ie., as phenomena to be studied in their own right (eg. Button, 1991b: ch.5-6).

A documentary method

Ethnomethodologists do not make *formal distinctions* between the general and the particular. Rather, they refer to what Garfinkel calls the *documentary method of interpretation* (eg. Garfinkel, 1967, p.76-103), which they claim is an dispensable resource for any person to assess what is definitely going on in a given event or an instance of action. The method consists of seeing or treating an actual instance as a document of some recognised or presupposed pattern. Furthermore, the pattern is confirmed and legitimised because of an actual instance that occurs. This method is used by all persons in their reasoning about the world. So, what seems to be at the core of sociological argumentation is the fact that anyone can identify what appears to be the norm, what the general conditions are in a given setting, what an instance is an instance of, and so on, as matters of *common sense*.

Ethnomethodologists have also criticised sociological investigations for employing the familiar common-sense world as an unexplicated resource. Instead, they have taken a procedural approach to common sense in reasoning and practice, reminding of the fact that people have a practical rather than a theoretical interest in their constitutive work. But, as Ten Have has pointed out, any procedural study, ethnomethodological or not, is faced with a problem he describes as the `invisibility of common sense'. It is a double-faced problem, he explains, minimising the unexamined use of common sense while maximising the examinability of common sense (Ten Have, 2004, p.31-32).

There were a number of critiques directed at the explanations ethnomethodologists gave

of the documentary method in the late 1960s, all of which can be said to be attempts to assimilate ethnomethodology with conventional method and reasoning. Firstly, the formal distinction between the particular and the general is raised where ethnomethodologists are accused of a phenomenological bias that forbids them to treat the relationship *between* individual descriptions and larger social bodies as formal (see Coleman, 1968, p.128; also cited in Denzin, 1969). Secondly, the assumption is made that the documentary method of interpretation must have a *nature*. This is evident in the complaint that ethnomethodologists fail to precisely indicate that nature, ie., what the method *is* independently of any actual interpretation. It is also assumed that meanings and definitions have *sources* external to the particular settings in which meanings and definitions are enacted and realised. As Denzin puts it, ethnomethodologists fail "to clearly indicate the source of meanings and definitions. In addition, they as yet offer no firm strategies for measuring the interaction process." (Denzin, 1969, p.929).

None of these critiques are particularly surprising. As I have already dealt with in the previous two chapters, studies in ethnomethodology continue to be misconstrued in the sense that ethnomethodologists draw attention to matters that are missed while they are critiqued for matters of no consequence for their enquiries. Ethnomethodologists have, in a studied fashion, steered clear of the tensions between the particular and the general, the tensions between topics versus resources of enquiry, of maintaining strangeness rather than going native, of questions regarding adequate measurements, warranted inference, idealised distance of a critical observer, and more, by avoiding altogether the quest for a `deeper' understanding. These are all problems for conventional methodological considerations, however, to be settled somehow in a satisfactory manner. But they become irrelevant if the unique adequacy requirement is correctly understood and met to the best of the enquirer's ability. This relates back to the argument I have made already, that quests for social-scientific understandings effectively miss the phenomena ethnomethodologists are interested in and draw attention to, and thus also miss the complex *surface texture* of what is going on in any

given setting (Lynch and Bogen, 1996). As regards the `ethnomethodological talk' of common sense in relation to enquiry, Sharrock and Anderson explain, that it "merely intends the fact that *amongst any given collection of persons organised into anything that can meaningfully be called a collectivity*, there will be a corpus of matters which those persons find `obvious' and `going without saying' as `beyond doubt and investigation'" (Sharrock and Anderson, 1991, p.63 [original emphasis]). In other words, for all practical purposes, we count on ourselves and others, *looking* and *seeing*, as the usual level of reliability. It is a `natural attitude', manifested in the way people react to and treat situations, the fluent way in which they go about most of their activities without having to hesitate all that much. One has no escape from *common sense*, thus understood, if one is an enquirer of the selfsame *common-sense* practice.

In avoiding the quest for a deeper understanding, ethnomethodologists choose to leave, as they are found in the world already, the distinctions between common sense and expertise. They are indifferent to whether or not there is anything epistemologically or principally different about the knowledge and results produced in all the occupational or other capacities people ordinarily operate in. This indifference should not be understood, however, to resemble the work in social epistemology which has accounted for both scientific and common sense reasoning as cognitively adequate (eg. Collins and Pinch, 1998). The adequacy of common sense arguments points to a form of relativism since common sense arguments can conflict with scientific arguments, for example, when you are told that the sturdy chair you sit on is indeed not solid at all, ie., at the atomic level (see eg. Ryle, 1954: ch.V-VI on this issue). But ethnomethodologists are not taking views on *relativism* or *positivism*. They are not taking views on whether social or physical realities are *objective* or *subjective*. They are not privileging natural science and mathematics as *more* objective or, say, *principally objective*, although the discussion above on Benson and Hughes's treatment may indicate otherwise (Benson and Hughes, 1991). Neither are ethnomethodologists in their enquiries negating objectivity as we ordinarily understand or experience, say solidity or

gravity. What they have had to say about these distinctions in relation to method and reasoning is to draw attention to the *fact* that `objectivity', `subjectivity', `relativism' or `positivism' are laboured and debated. The disappointment here, for the idealised observer, is that ethnomethodology cannot end a dispute on a controversial subject. It does not provide an authoritative basis for normative judgements that can override the prejudices of common sense. The idea is to use ordinary modes of observing, questioning, describing or comparing, all of which should be expressed in plain ordinary language—in a word, a `normal science'. In addition, when showing what experts say and do, only their vocabulary should be used rather than a vocabulary of a theoretical framework for *interpreting* what is really going on in what the experts say and do. As Lynch puts it, "[t]his normal science offers an analysis that is thoroughly "contaminated" by native intuition, vernacular categories, and commonsense judgements." (Lynch, 1993, p.305). I will now discuss policies for selecting and analysing cases.

Searching for primitive cases

In Chapter 7 of his book, `Scientific practice and ordinary action', Lynch lays out a list of policies for studying scientific practice. The study policies of ethnomethodologists are not entirely in unison but I have given consideration to the policies on Lynch's list with respect to the development of this thesis, and some are already addressed in this chapter. I have discussed a unique adequacy requirement which shifts from the instructions to recover the `core activities' of a practice (eg. Garfinkel, 1967: ch.8) to the singular demands of a `unique adequacy' way of coming to terms with perspicuous settings. In other words, a description must enable an adequate reproduction of an intuitively recognisable action while providing a *notational index* of the details of the action's performance. I have discussed ethnomethodological indifference which avoids attributing a special epistemological status to so-called *lebenswelt* pairs in expert practices (eg. Garfinkel et al, 1981). This is not to say that hackers like the operators on the archive cannot be recognised as having singular hands-

on engineering knowledge about what they do and how they do it, what the general and particular characteristics are of the work they do and the objects they handle. It is simply that an ethnomethodological enquiry of the work taking place in daily administration and use of the archive will not secure a method by analogy with some scientific or engineering expertise, not that put to work on the premises nor any other. The method suggested instead is to use the normal science methodology suggested by Lynch (Lynch, 1993, p304-305), and be sensitive to naturally occurring scenarios and natural attitudes in drawing attention to the surface textures of practical action and reasoning. I will use this section to discuss the policies for selecting cases for enquiry and, in doing that, I continue to refer to Lynch's list.

Lynch suggests that enquirers should search for perspicuous examples of the practice of interest—examples in which the topics unique to that practice have a *vernacular role* as he puts it (Lynch, 1993, p300-302). Enquirers should then investigate actual cases of such examples, ie., search for naturally occurring primitive scenarios and investigate them in detail. The idea here is to search for recurrent, routinely enacted, familiar, observable and comparable scenarios although no single case can formally stand in for all the others which are similar (Lynch, 1993, p.300-302). We shall then consider these suggestions in relation to the question of *what to look for* and *what to do next* in the daily operation and use of the archive.

The archive ran twenty-four hours a day, seven days a week, every single day of the year. Background operations were never idle and new preprints could be submitted to the system any time, day or night. There were internal time cycles, however, corresponding to the mundane working hours of staff over five days per week. The 4 pm weekday *freeze* was the delimiter. All submissions since the last *freeze* would be cleared for public dissemination.

My accumulated log reaches over 450 weekdays with anywhere from 22 - 238 new preprints routed through and cleared for public access on each of these days. This does not include the numbers for incoming new versions of existing preprints, cross-listings between

subject archives, journal references added, and more. It is obvious that the magnitude of data is overwhelming but useful still in their entirety to produce simple figures as I just did. Simple figures can indeed substantiate a few claims. Consider the following dialogue:

Editor:	Look at this! This page alone has 5 typographical errors. How do you explain that?
Proofreader:	I'm very sorry. These things occasionally slip through. It is unfortunate but we are trying our best.
Editor:	Trying your best! Do you have any idea how annoying it is for the reader to stumble across 5 typographical errors on a single page?
Proofreader:	Well, I can see that of course. Allow me to point out though - this page, for example, has 612 words so 5 words with typographical errors are not even a single percentage, just about 8 per mill actually. And, if you look at it from the perspective of typography - see, the page has 3677 characters and only 12 characters need shifting as far as I can tell. That's roughly 3 per mill. Really not too bad.
Editor:	Do you think we're not serious about this publishing business? Either you people are proofreaders or you're not.

In this fictional conversation, a proofreader is providing statistics rather than recognising that typographical errors are a problem. The outliers are so few, statistically, that they should be nothing to worry about. Although a perfectly sound statistical argument, it is absurd in relation to the question, “[d]o you have any idea how annoying it is for the reader [...] ?” or the statement about the seriousness of a publishing business. Proofreading is finding all typographical errors and acting on them. It is a matter of producing a complete and consistent product. The presence of typographical errors draws attention to a `flawed' product, but it also draws attention to the meticulous work of inspecting, detecting and (re)arranging types. The analogy with the daily administration of preprints on the archive is simple, ie., the nuts and bolts of it is coping with outliers in order *not* to have a flawed product on the archive website.

Two somewhat confusing issues are raised in this thesis with respect to the policies of selecting cases. I have already discussed to some extent matters of *order*, the *mundane* and *business-as-usual*. Ethnomethodologists are concerned with the *social order* and suggest that order can be respecified in and as of any naturally occurring primitive scenario. Another

interrelated issue in now raised with respect to the policy of searching for naturally occurring primitive scenarios that are recurrent, routinely enacted and familiar—how the mundane, business-as-usual is achieved. As regards the matter of *order*, what risks being confusing is that the actual labour communicated in the operation and use of the archive centres on matters of disorder, of keeping things on an even keel in the face of broken objects, spurious actions, fragmentation of method, variations in solving particular problems, and so on. If the statistical data are consulted on, say, successful processing of `TeX`, this automated self-archival service takes care of itself, more or less. The average number of new preprints in 2002 where close to 140 per day. The average number of problems communicated and solved specifically because of `TeX` processing failures were only 4-5 per day. That comes to about 3 percent. Similarly, problems communicated about oversize figures produced with the `Adobe Illustrator` software in the year 2002 were less than 30 out of roughly 36 thousand new preprints submitted during that same year.[7] In a word, most things were in order at the end of the day, the usual business of a more or less automated self-service. But, as I have also explained, outliers could not be ignored. So, what further risks being confusing is that the actual work that needs doing is focused on the *unusual*. The mundane is expecting the *unexpected*. The recurrent, routinely enacted is indeed a practice of *hesitation*, of *double-checking* and of *suspecting* that all things at hand are possibly not what they should be. In other words, the familiar is somewhat confused.

What I will do, for the present purposes, is to assume that there are orders of disorders which *can* be shown, ie., their respective achievements if they are adequately examined. I will also assume that doing the work that needs doing in the operation and use of the archive can only be achieved in the absence of systematic doubt such as whether digital infrastructures, objects, instruments, code, documents or codes of conduct, are what they are. In that respect there are mundane everyday procedures to enquire into. However, coping with the selfsame work can only be achieved by assuming that system components can at any

[7] These numbers representing processing failures and oversize figures are estimates from probing, +/- 10%.

point fail to keep up with new material. The digital objects, supplied by submitters, can always be disorderly and thus they need meticulous checking. To summarise, I will assume there are orderly, mundane and business-as-usual practices of coping with the unusual and the unexpected, of hesitating and double-checking.

To clarify, this thesis did not set out with an ambitious agenda of sorting out foundational issues of order and disorder, the familiar and the strange. It is certainly not making any foundationalist claims. It set out with a modest agenda, of simply showing how in the operation and use of the archive, the people involved make the work they do visible-accountable. I suggest that the radical phenomena of doing this work, are phenomena available to the policies of ethnomethodology in the sense that if they are enquired into, procedurally and thematically, questions such as what the specific restrictions are and what the responses are, *can* begin to be answered and matters of *social relevance* clarified. The phenomena are found in the work of *inspection and detection*, *corrective action*, and the *use of tools* to accomplish that work.

Summary and discussion

In this chapter, I have reviewed a number of methodological issues immediately relevant for the present purposes—questions raised about research and research sites, data collection and data analysis. I have drawn attention to common considerations in sociology such as securing an observer's *point of view* and asking whether there is not a *stricter* and *more demanding* level of enquiry than using common sense to get at an understanding of social phenomena. I have described and explained the basic arrangements for operating and using the archive, my own participation on board the daily administration of preprints, the recording of data, and a policy for selecting and analysing cases. In doing that, I have also drawn attention to critiques raised by ethnomethodologists about social-scientific methods and reasoning, for instance, how the familiar common-sense world is employed as an unexplicated resource, while it is made difficult, if not impossible, to perceive and

reconstruct lived social phenomena prior to the adoption of theoretical and methodological prejudices. I will now summarise the main points and further discuss the implications of the study policies I have adopted as opposed to standardised methods.

Two things, specific to the arrangements for operating and using the archive, are consequential for the ways in which data are collected and cases to analyse made relevant. One concerns the fact that this mass-dissemination system was set up to take care of itself, insofar as that was possible. It means that the actual labour of attending to data and document processing is one of detecting and coping with that which cannot be resolved without the involvement of the operators. The other concerns the actual place where the work that needed doing on the day was communicated, and the reporting of goings-on made visible, ie., the message box. Consequently, the data for consideration in this thesis are recordings from this box and the cases chosen for analysis are perspicuous settings of detecting and coping with outliers and anomalies.

A number of arguments can be made to the effect that this enquiry lends itself to conventional ethnographic concerns and considerations. For example, this could be considered a conventional case study that draws boundaries around a particular workplace and its immediate surroundings (Stake, 2000). There are well known concerns raised about research sites being prominently virtual (Jones, 1999; Hine, 2000), about the use of documents (Silverman, 2006; Scott, 1990), or to embark on prolonged on-site exploration of an `esoteric domain' (Hess, 2001). Hess points out, for instance, that researchers of specialised practices become accustomed to the use of a mixture of methods and perhaps working with a rather confusing concept of *fieldwork*. They hang out in laboratories, talk to experts, take notes and set up audio or video recorders. Some develop a comprehensive understanding of a science or a specialised know-how over time. Some might have a background in science or a technical training, are hired by those they study or they work side-by-side with them as I did while I was on board the daily administration of preprints on the archive.

This chapter has dealt with these and other common considerations in the sense that I have clarified the ways in which they are avoided or bypassed when the study policies of ethnomethodology are adopted. Methodological themes that are typically associated with sociological research will underscore issues of appropriate application and what pitfalls should be avoided. They formulate propositions and articulate common problems, centring on the "use of criteria, decision rules and models which tie research designs to scholarly traditions." (Lynch, 1999, p. 212). As I have elaborated, there are problems with criteria, decision rules and models which are difficult to resolve. For example, measurements are always indirect, empirical references inadequate, and law-like statements not adding anything to what is already observable in the orderliness of a site of activity. There is no escape either from *common-sense reasoning* about social phenomena even if the objective of sociological investigations is to develop a credible *sociological reasoning* about social phenomena. Furthermore, ethnomethodologists have expressed uneasiness about interpretivism in the storytelling of ethnographers which is built around presupposed characterisations of phenomena, cognitive maps and related devices (eg. Lynch, 1993).

Sociologists accepted many of the criticisms that were first voiced in the 1960s. As Denzin (1969) puts it, the thrust of Garfinkel's critique is that,

> [t]he sociologist has pursued his normative system at the expense of concrete behavioural analyses of face-to-face interaction. By forgetting that he is responsive to social demands, the sociologist has overlooked the irrational elements of his own conduct–most notably his inability to make sound observations, to reliably code documents, to conduct face-to-face interviews and on. (Denzin, 1969, p.932-933).

Interestingly, the responses to these criticisms have centred on improving criteria, decision rules and models—to develop better schemes for observation, the coding of data and rigorous analyses, and to use `reflexive accounting' which acknowledges the shortcoming of both quantitative and qualitative methods (eg. Denzin and Lincoln, 2000; Atkinson et al, 2001). What has not happened is mass-uptake of ethnomethodological study policies. Rather, they tend to be misconstrued as instructions on how to improve conventional methods or

otherwise they remain on the margins of mainstream sociology which cannot accept that no principle will sort out conflicts—that actions and events cannot be adequately identified independently of the order in which they are embedded or that order which is itself the site of actions cannot be identified independently of those actions. In sum, to abandon the theoretical leverage provided by an idealisation of a scientific observer, using established standardised methods, seems unthinkable for most sociologists.

Although ethnomethodology took, as its starting point, inspirations from phenomenology and figures like Alfred Schütz (see eg. Schütz, 1943), the link with Schütz and phenomenology does not mean that ethnomethodology is radically subjective, for instance, that it derives fundamental properties of actors' experiences from properties of *pure consciousness*. For the purposes of meaningful enquiry, ethnomethodologists challenge *subjectivism* which takes as its starting point that anything goes and that perhaps method should be voted on (Feyerabend, 1975). Similarly, they challenge *objectivism* in the natural sciences which takes as its starting point an unquestionably given world *out there*. They also challenge sociology's unquestioned faith in the existence of social facts and lack of adequate reference to the ways in which such facts are achieved. So, the phenomenological bias, mentioned for example by Coleman (Coleman, 1968), is not preventing ethnomethodologist from doing what necessarily needs doing in order to make credible inferences and solve inductive problems. Ethnomethodologists are not solving the inductive or inferential problems of an established social science and they are not taking views on the labouring and debating of *subjectivism* or *objectivism*, as such activities can be found in the world.

Sociological investigations have inherited methods from the empirical sciences as well as philosophical scepticism from Descartes (Descartes, 1991[1637]; also discussion in Sharrock and Anderson, 1991). The sceptical attitude essentially relates to a distinctive form of enquiry, as Sharrock and Anderson point out, a `pure' enquiry of the `pure' reason which establishes `true' knowledge. Philosophical scepticism thrives in the idea that judgements about the nature of phenomena or the relationships between them need a special

epistemological warrant in relation to empirical evidence. But the question of whether there is anything principally or epistemologically different about knowledge or the results produced by lay persons in society, a sociologist or some other professional/scientific expertise, is one that ethnomethodologists remain indifferent to (Lynch, 1993: ch.7). Philosophical scepticism also thrives among sociologists who subscribe to rule scepticism and the perceived indeterminacy problem of interpretation, both of which have been underpinned by a particular reading of Wittgenstein's later work (Wittgenstein, 1953). Wittgenstein argues that we should not ask for *meaning* but *use* which has been understood to mean that `use', *as distinguished from* meaning or the rule being used, refers to social practices, a *form of life*, which cracks the `mystery' of indeterminacy and makes possible decision-making, rule-following and consensus about truths (eg. Bloor, 1976; Collins, 1990; Collins, 1992). This reading of Wittgenstein is rejected by ethnomethodologists (eg. Lynch, 1992b; Bloor, 1992; Lynch, 1992a; Sharrock and Button, 1999) and treated as another example of sociologists needing to hypothesise stable external arrangements—external to observable orderliness. The counterargument is that the *problem of interpretation* or *rule-following* should be treated as *completely internal* to the site of interpreting and rule-following. But indeterminacy is seen as a theoretical or inductive problem for which sociologists have attempted to find a remedy.

In light of these considerations, it remains to be said that unified methodological protocols and descriptive models are missing from the programme of ethnomethodology. The study policies also vary and have changed over time, as this and previous chapters indicate. One way of looking at it is to follow Lynch's directive and treat methods as discoverable (Lynch, 1999) and to assume that distinct topics of order or disorder can be found in the details of lived action and interaction—topics which give rise to coherence, as Lynch puts it (p.218). But, although the emphasis is on a set of study policies rather than uniform protocols, ethnomethodologists still have a recognisable style of reasoning which separates their enquiries from those that do not adopt the study policies of ethnomethodology. In spite

of the policy of a normal science and the suggestion that enquirers should use common everyday language and the language used by those whose actions they study, ethnomethodologists are not exempt from styles of reasoning that bring new types of objects into being, new evidence and new possibilities. As Ian Hacking puts it, these are the novelties created by scholarly reasoning and concerns the ways in which objectivity comes into being —how analyses are given a name, a sense of purpose and identity (Hacking, 2002). One can argue, however, with good reason, that anyone's style of reasoning concerns the ways in which objectivity comes into being in a transparently recognisable and thematic fashion. For the present purposes then, style will have a double function. It concerns the natural absence of systematic doubt so that the people involved in the operation and use of the archive – as well as the enquirer – can proceed in ways that are to them intuitively recognisable and thematic events-in-a-social-order. It also concerns the project of *entering into the phenomenal field of operating and using the archive*—perceiving and reconstructing occurrences so the reader can see what is said and done in ways that are intuitively recognisable and transparent, as well as thematic.

Chapter 5

Inspection and detection

Introduction

This chapter will focus on the work of inspecting the submissions when they were uploaded onto the archive. As I have already mentioned, submissions comprised of bibliographic data and preprint files. The former, often referred to as *metadata* or simply *abstracts*, contained author names and affiliations, title, abstract, classification codes, etc. The actual preprint files, however, typically included one or more `TeX` files, auxiliary formatting files for `TeX`, and figures or photos in `postscript`, `GIF`, `PNG` or `JPG` formats. Occasionally, a single `PDF` or `postscript` (`PS`) was uploaded. This was the case when authors did not use `TeX` or when they were uploading an export from their own `TeX` processing environment.

I will take examples of how the work of inspecting the incoming packages was made possible, who/what was inspecting them and what new submissions were inspected for. A considerable amount of this work was delegated to software applications with manual oversight and interception. One can argue that the operators negotiated the finer and courser grains of inspection with the devices they had gradually built to support this work. The courser grains were delegated to computational harvesting and diagnosis, while the finer judgements were made by operators consulting documentary evidence of anomalies in reports, logs and the responses that certain scripts would display.

I will begin by discussing two auto-responses to users of the archive. The first was received upon successful registration with the system as a submitter and the second upon successful upload of a preprint. These responses draw attention to the work that was expected

of registered submitters, ie., what their responsibilities were and what to expect or not expect from the operators. Then, I will explain the tools available to the operators to inspect reports, objects and messages that reveal the presence of anomalies. Thereafter, four sections consult and reconstruct common scenarios in the daily grind, each of which involve inspection work. The final section of the chapter will discuss these scenarios in reference to their relevance to the central concerns of this thesis.

Being a submitter

Registering as a submitter of preprints to the archive was much the same as subscribing to a conventional mailing list, except that subscriptions had to come from an email account at an academic or otherwise `legitimate' research facility. Some personal details were required such as full name and the area of research to which the new member had plans to contribute. Upon successful registration, a pre-written email response was generated by the system. I quote the first lines:

Fragment 5:1

```
1   ==========================================================================
2
3                The [...] e-print archive is fully automated.
4
5              It processes over 150 new submissions per day.
6
7   This is only possible if YOU as author or submitter take responsibility:
8   always carefully check and verify your submissions, pay close attention to
9   diagnostic messages sent to you, and take corrective action if necessary.
```

The newly registered submitter learns from the top of this message that the archive is `"fully automated"` and `"processes over 150 new submissions per day"` (lines 3 and 5), although, in lines 7-9 it is made clear that the responsibility for this process is up to `"YOU"`. Full automation is `"only possible"` by following the imperatives to `"carefully check [...] verify [...] pay close attention [...] and take corrective action"`. In other words, full automation is conditional upon new submitters being vigilant and responding *correctly* to future diagnostic messages.

The following pre-written email response was generated upon successful submission of a preprint:

Fragment 5:2

```
 1 ------------------------------------------------------------------------------
 2 Successful submission to the archives can be a significant source of pride and
 3 accomplishment. It entails many serious responsibilities: if you cannot
 4 check off on all of the items below, then you should replace your submission.
 5 (Note: *do not* resubmit, instead use *replace*)
 6
 7    ___ The Title/Author fields above are correct.
 8    ___ The abstract is complete, correct, and wrapped correctly.
 9    ___ Capitalization in title correct (we automatically lower case titles with
10        excessive use of upper case, but we get some acronyms incorrect
11        --replace with uppercase letters only where they should be).
12    ___ The authors are listed in `Firstname Lastname' order.
13    ___ any anonymous ftp or http pointers to additional files are given in
14        standard url format (e.g. ftp://myhost.domain/path/filename.ext or
15        http://mywwwhost.domain/path/filename.html)
16    ___ Periods are separated by a space from the end of any URL's
17    ___ MSC-class:, if included, can be a single AMS subject classification, or
18        can include several, in the form `### (Primary) ### (Secondary)'
19    ___ Any Journal-ref is a complete bibliographic reference to an already
20        published version (includes volume and page number info in the case of a
21        print journal).
22    ___ No raw (uncompressed, un-uuencoded) PostScript was submitted (by e-mail)
23    ___ All files including PostScript figures were included.
24    ___ This paper has not been `put' to any other […] e-print archives.
25    ___ The full text of the paper itself is available directly from the
26        archive.
27    ___ The submission can be retrieved uncorrupted (TRY IT).
28
29 Submitters who repeatedly leave errors uncorrected may lose submission
30 privilege to the archives. (Replace using the current tentative papernumber
31 unless notified otherwise.)
32
33 […]
34
35 For further hints, keep reading or get the help text by sending a message
36 with the subject `help', or see http://archive.domain/help/ . Guidelines for
37 a properly formatted abstract are given in prep.txt (available via the `get'
38 command or at http://archive.domain/help/prep ). If you have not already read
39 this, please read it now to ensure that you have followed the guidelines. It
40 is an abuse of valuable archive admin time to edit abstracts by hand, and
41 negligence will be duly noted and could result in prejudial future treatment.
42 […]
```

A number of different strategies are used here to recruit the submitter in order to reduce incorrect submissions. The submitter is both *warned* of `“many serious responsibilities”` and *encouraged* that `“successful submission”` is a `“significant pride and accomplishment”` (lines 2-3). This is followed by a check list of thirteen items that meticulously *instruct* the author on how to avoid incorrect submissions (lines 7-27).

The first eight items on this list refer to metadata. They should conform to

"[g]uidelines for a properly formatted abstract", a task which is familiar to cataloguers who quality check records. The last five items on the list, however (lines 22-27), spell out possible errors with the preprint source files, errors that may well occur even if they have been processed without a halt. Submitters are explicitly told to eliminate missing files and to ensure that there are no double uploads. The preprint should also be retrievable uncorrupted from the archive website. Checking off this last item may call for expert knowledge of how to prepare TeX, and how embedded files such as the auxiliary format files or how figures in postscript format are interpreted by TeX compilers.

The email response is not only concerned with facilitating the responsibility of the submitter but also *preventing* submission anomalies. The text that follows this check list warns submitters of leaving errors uncorrected. It refers them to guidelines and warns them that if submitters do not take "corrective action", operators may have to "edit abstracts by hand", which is an "abuse of valuable admin time" (line 40). There is also an overhanging penalty for failing to take responsibility. Authors "may lose submission privilege" and "negligence will be duly noted and could result in prejudi[c]al future treatment" (lines 29-30 and 41).

As I explained in the previous chapter, submitters were geographically distributed. To show how submission procedures could unfold on their end was not available for the purposes of this thesis. But I could do it myself. There were three protocols for submitting a preprint, a mail protocol, a file transfer protocol (FTP) and a web page protocol, ie., graphical interface on the archive website. For the submission of metadata in particular, the mail and FTP options required that a transmission of a complete text string would come out correctly on the other end or the whole thing would have to be uploaded again. The web submission however, was designed so that inserting metadata into the relevant fields would trigger some auto-checks at the front end. But the interface did not neatly restrict and confine choices by its layout and structure. It did not enforce a narrow range of options at auto-

checking and self-checking points. If that had been the case, many more errors, such as indicated in the check list, could have been eliminated at the front-end. In other words, the first eight items on the list in fragment 5:2 concern the task of inspecting *already uploaded* metadata.

The situation with the preprint files was somewhat different. Many things could go wrong in the submission process. Packages exceeding a certain size limit were automatically rejected and thus not uploaded successfully but dumped into a temporary holding area, available to the operators, while the submitter received an auto-response explaining what had happened and what to do about it. The same would happen with `badly' configured `TeX`, certain combinations of files and formats, and the single-file `PDF` or `postscript` submissions if they could be auto-diagnosed as `TeX` exports. In other words, computerised inspection and detection would result in a host of auto-rejections and the submission would not go online. However, auto-rejection could not always be safely implemented. Thus, there were examples of successful processing through the `TeX` auto-compiler although the `postscript` output turned out to be corrupted and failed the final data integrity check. Hyperlinking features might be missing or not working, figures could be out of place, and more.

Matters of social relevance

In the last section, we could see in the fragments of auto-response messages that *practical responsibility* and *guidelines* are matters of relevance in using and operating the archive. An issue of responsibility is already manifested in the two messages quoted in the previous section. *YOU* are told to *carefully check*, *verify*, *pay close attention* and *take corrective action*, and `“negligence will be duly noted”`. Matters of legitimacy are manifested in *fields*, *wrapped abstract*, *upper case*, *lower case*, *acronyms*, *`Firstname Lastname' order*, *pointers*, *MSC-class*, and other specifically named phenomena which need to be inspected and checked off as *correct*, *complete*, *in order*, *standard*, *classified* and

uncorrupted—in short, according to guidelines.

Substantive evidence could be detected that submitters were inspecting and correcting their packages. A temporary record of repeated attempts to submit was available and the differences between each attempt observable. Messages from submitters about problems communicated where they had looked, what they had observed, and what they had done or not done. A host of guidelines was written in help pages on the archive website, as pointed out in the last paragraph in fragment 5:2. Preferred package/document types were listed on these pages and explanations given as to what is acceptable and what will or will not work. It was also stated clearly what help to expect, or not to expect, from the operators on the archive. The email response to new subscriptions, quoted in fragment 5:1, continues as follows:

Fragment 5:3

```
 1  There is no secretarial staff to manually correct mistakes, fix typos, amend
 2  layout, or perform other remedial tasks. In particular, there is no one to
 3  guide submitters step-by-step through simple submission procedures explained
 4  at length in the help texts, nor are there resources to assist with generic
 5  problems of word processing software or packaging of submissions. Use of the
 6  e-print archive is free of charge, and this is feasible with a skeletal staff
 7  here only insofar as users take full responsibility for their submissions.
 8
 9  Staff time here is dedicated to improving the software and adding features,
10  and tuning the server and mirror network, rather than assisting individual
11  users with minor problems that can be solved entirely at the user end.
```

In other words, new members were made aware that something so *skeletal*, in terms of staff on the archive, could not disseminate 150 new preprints each day if the operators were expected to take `"corrective action"` every time there was some error. Resources were scarce and the service was `"free of charge"` (line 6). The operators were not `"secretarial staff"` (line 1), engaged in customer service, and thus did not attend to `"remedial tasks"` (line 2) that could be solved by submitters themselves. The point is made very clear:

```
There is no secretarial staff [… to] perform […] remedial tasks […] no
one to guide […] through simple submission procedures explained at
length [elsewhere … only] skeletal staff […] improving the software and
adding features.
```

One can argue that the incentive to carefully inspect and correct the metadata should be that precise and consistent data means more robust searching and browsing on the archive. This is not explicitly stated in fragment 5:2 but submitters, as experts of data and computer-intensive disciplines, should know that the visibility of authors and their works on the archive hinges, in part, on orderly metadata. As it turned out though, the metadata were sometimes left in a state of poor quality by any records management standards. Similarly, one can argue that the incentive to carefully inspect and correct preprints should be that a corrupted preprint does not view correctly, if it views at all. It might require from an interested party that they download the source files and figure out how to process them successfully on their own computers. But even if submitters were receiving notices, for example, flagging a failed `postscript` integrity check in boldface, `` `*** PS BAD ***' ``, some of them did not act on the warning or they claimed not to understand what to do in their communication with the operators. As we shall then see, the matter of practical responsibility, as one of social relevance and consequence is, in the main, manifested in the work of inspection and detection taking place on the back end in response to the fact that submitters could not be relied on.

On the matter of guidelines, there were some *rules of the game* to consider. Anyone from a recognised institutional account could contribute to one or more of the subject archives as long as their preprints were `topic relevant' as the operators put it. Subject archives had moderators not reviewers. Bibliographic data entered the environment more or less unchecked, as I have said, and preprint files only had to agree with the auto-compiler to go through successfully. But the general idea was that metadata should have the same or similar representational formats with no attempts to draw attention to any one of the preprints in the listings. Registered submitters should not upload packages for third parties. Each package should be a self-contained scientific article including all the figures and other support materials it referred to. It should be legible, using scalable fonts and `reasonably' sized figures, ie., efficiently produced. It should also be uncorrupted, as we have discussed, and not be a book advertisement, an incomplete report or errata to a previous preprint, and there were

non-negotiable copyright issues. The list goes on but these are the conspicuous examples.

The operators on the archive had learnt from experience that submitters and their submissions were not always in order. However, there was no way of foreseeing exactly *what* was not in order so they systematically inspected for anomalies on the back end. Namely, achieving the stated, the `“e-print archive is fully automated”`, hinged on responsible conduct at the front end or, otherwise, interception and repair on the back end. Metadata might not be according to guidelines, preprints might be incomplete, illegible or somehow not legitimate, and processing problems could always occur. In other words, the archive was not an *automata* or *fully automated* in any strong sense. The system was configured to warn submitters about what they needed to do and what to avoid, indicating the lengths to which operators would go in order to minimise service labour on their end and leave all the efforts to submitters. What we have seen in fragments 5:1, 5:2 and 5:3 is that particular activities are named as those manifesting responsibility. A number of phenomena are also named specifically as those which need to be inspected and checked off. What is not evident in these messages, however, are the precise manifestations of that which, on inspection, is detected as *not* being those things—*that* which is *irresponsible*, *incorrect*, *incomplete* or *corrupted* and thus of prime interest in the administration of preprints as phenomena that *can* be seen and intercepted.

Tools for inspection and detection

We have established that inspecting for and detecting anomalies was necessary on the back end. The operators corrected errors and mistakes, and worked out inventive ways around a range of problems. I will enquire into *taking corrective action* in the next chapter, but in this section, I will introduce the devices and controls that were available to the operators in order to look for and see anomalies.

An operator would be using a minimum of four terminal windows - **`A`**, **`B`**, **`C`** and **`D`** - the

main purposes of which can be summarised as follows:[8]

Fragment 5:4

```
 1 A- News reader
 2 Metadata reporting on each of the incoming packages are routed into a news
 3 reader. They are read through by an operator, one by one, to catch anomalies,
 4 some of which the system does not or cannot flag specifically.
 5
 6 B- Mail reader
 7 All incoming email to admins and internally auto-generated messages go to the
 8 mail reader. Contents include:
 9 * Auto-reporting on incoming submissions and machine-detected problems
10 * submitters asking for help with file formatting and TeX styles
11 * submitters asking for help with figures and photographs
12 * submitters requesting special dispensation because of too large submissions
13 * submitters complaining about preprints being moved between subject archives
14 * submitters complaining about preprints being removed
15 * submitters requesting lost paper passwords
16 * operators communicating about submissions and/or submitters
17 * operators communicating with submitters
18 * operators communicating with moderators and miscellaneous others
19
20 C- Local terminal
21 On a local terminal, the operators can inspect for problems with submissions
22 (rejected or not) and work on fixing them by copying the submission over from
23 the remote server using the `hack' script. This might be necessary because of:
24 * Postscript image format failures
25 * TeX definitions failures
26 * necessary editing/alterations of images and/or other files
27 …and then they copy the submission back to the main server if fixing is
28 possible.
29
30 D- Terminal access to the main server
31 Operators execute various commands/scripts on the main server in order to:
32 * remove new submissions using the `lastmv' or `lastrm' scripts
33 * remove new versions (another day replacements) manually
34 * run “autofixes”, a script that suggests corrections to bibliographic data
35 * edit directly bibliographic data using command-line text editor
36 * manually process submissions and new registrations
37 * schedule manual changes for mirroring
```

These were the necessary terminals to have up and running. Other necessary devices were `postscript` and `PDF` viewers and a web browser in order to access and inspect fully processed preprints online before they were cleared for public dissemination.

As I have already mentioned, the application software on the archive server inspected and diagnosed submissions packages at the entry point of each upload. The bibliographic data and the preprint files were routinely diagnosed. Data relations and dependencies, associated with each submission, were generated. To summarise, these processes accumulated *awareness data* about each package and distributed that data to databases inside the server machine so they could serve automated and semi-automated processing later down the line.

[8] This overview of terminal windows was prepared specifically for senior library staff.

Standardised messages were generated and emailed to submitters as well as warnings if anomalies were detected and an error log would be sent if TeX compilation failed. Reports with a selection of data on each new submission were also routed into the news reader (fragm. 5:4, "A" lines 1-4). They would line up, one report per incoming package in chronological order, grouped under the headings of each subject archive [see Appendix III for a list of subject archives in the system]. In this way, diagnostic data and metadata provided crucial access to new submissions.

There was an agreement about a weekly rotation as to who was the `principal actor', an operator who systematically inspected this queue in the news reader, who also responded to diagnostic overviews of submissions on the day, and responded to requests and complaints from submitters. One or two other operators would `lurk' in the system. They monitored, `listened in' on the exchange, and pitched in with suggestions, tips and hands-on work as needed. Designating a principal actor is one way of avoiding double efforts and the risk of confusion over who does what. Only one person at a time can give a particular command and because the operators did not literally see or hear each other, they relied on the fact that they could all see what needed checking. They could all see alerts that came into the shared message box (fragm. 5:4, "B" lines 6-18), as well as the communication with submitters about their packages. They also communicated among themselves in reference to particular anomalies and warnings which were detected in the course of inspecting messages, reports and preprint files. They each had direct access to the incoming packages, either by loading them onto their own desktop machines via terminal access to the main server (fragm. 5:4, "C" lines 20-28) or by accessing them directly on the main server (fragm. 5:4, "D" lines 30-36). I shall now walk through examples of inspecting reports, preprints, and the messages box.

Inspecting metadata reports

What I will dwell on throughout the next four sections are the actual indicators that become subject to administrative scrutiny, phenomena acting as reminders or indicators of potential as well as known disorders. I will begin by showing a system-generated report of a submission as it would appear in my news reader.

Fragment 5:5

```
 1 ------------------------------------------------------------------------------
 2 \\
 3 Paper: archive/papernum
 4 From: full name <username@address>
 5 Date: Mon, 11 Nov 2002 09:53:23 GMT (33kb)
 6
 7 Title: Title
 8 Authors: Firstname Lastname
 9 Comments: 24 pages, 3 figures, RevTeX4
10 \\
11  abstract
12 \\
13
14 Contains:
15  figure-1.eps: 7407 bytes
16  figure-2.eps: 9425 bytes
17  paper-file.tex: 107366 bytes
18  figure-3.eps: 13784 bytes
19
20 Stored as: papernum.tar.gz (33kb)
21
22 Author Warnings:
23
24 Author 1: Firstname Lastname
25
26  paper-file.tar: unpacked 4 files from tar file
27
28 PS files:
29  papernum.ps.gz (102kb)
30
31 To verify abstract and postscript, use
32 http://archive.domain/abs/archive/papernum
33   User: archive/papernum, Password: passwd (access still password
34 restricted)
35  Abstract will appear in mailing scheduled to begin at 20:00 Monday
36   US Eastern time (i.e., Tue 12 Nov 02 00:00:00 GMT).
37
38
39 Admin Warnings:
40 Uploaded 36148 of 36148 bytes from username at IP address in 0s
41
42 paper-file.tex unpcified
43 Mon Nov 11 05:53:26 2002: AutoTeX took 3s
44 gs check(s) took 3s
45
46 orig files held in /u0/incoming/puts/temporary-identifier
```

Lines 1-12 contain bibliographic data which were parsed into separate files, `papernum.abs', and then rendered onto the archive website as front pages for respective

preprints. An archive identity, "**archive/papernum**" was generated and reported in line 3. The full name and email address of the registered author/submitter is detected and reported on line 4. A date-time stamp is generated and reported on line 5 along with the size of the compressed submission package in parenthesis. The rest is data provided by the submitter: title, authors, a comment section, an abstract, and occasionally additional sections holding bibliographic reference and subject class codes (lines 7-12).

Lines 14-46 list further diagnosis. The contents of the package are listed on lines 15-18, reporting file formats and file sizes. The author(s) list, here on line 24, will show the author names as they were formatted by the submitter and at least one of them should be recognisable in the "**full name <username@address>**" on line 4, ie., the submitter. Line 26 reports how many files were unpacked from the package which should accord with the `Contains' section on lines 15-18 as well as the `Comments' section on line 9 if the submitter has listed the number of figures, here, "3 figures". The formats for figure files listed in the `Contains' section, should be indicated with PS, EPS, EPSI, JPG, JPEG, PNG or GIF extensions to their file names. Lines 28-29 report that a postscript has been generated from the preprint files and lines 31-36 display an excerpt from the auto-response sent back to the submitter about what to check, how to access it, and when public dissemination is scheduled. The remaining lines (39-46) report on the upload of the submission package. Line 40 reports how many bytes were uploaded, from where the package came, "**username** at **IP address**", and how long it took to upload. Line 43 reports exactly when the TeX auto-compiler was invoked and that it took three seconds, "3s", to generate the postscript output and line 44 reports that ghostscript, "gs", also took three seconds to check its data integrity. Finally, on line 46, the location of a temporary holding of the original package is reported in the unlikely event that irreversible damage is done to this submission before it is cleared for public dissemination.

An experienced operator will only glance for a moment at each of these reports in order

to target anomalies that should be attended to before the 4 pm freeze. But the example above does not contain any anomalies. The `Comments' section (line 9), naming 3 figures, matches the list of three ".eps" files in the `Contains' section (lines 15, 16 and 18). The author(s) list on line 24 shows Firstname Lastname format and the name corresponds to that of the submitter (line 3). The abstract, omitted here, would be inspected and diagnosed separately by a relevant script, `autofixes', and the output inspected by the principal actor later in the day. Finally, lines 40, 43 and 44 indicate nothing of notice. The total volume was uploaded and the TeX processing and integrity check completed swiftly. We shall now take some examples of anomalies.

The completeness of preprints was one of the main requirements. Preprints should be self-contained, ie., include all the figures and necessary TeX auxiliaries. The matter is mentioned in the auto-response to submitters (fragment 5:2), more specifically, item 12 on the check list which tells the submitter to check that "[a]ll files including PostScript figures were included". (fragm. 5:2, line 23). But, as experience had shown, packages could be incomplete in some way so the operators inspected for indicators or evidence of that being the case.

Fragment 5:6

```
 1 ------------------------------------------------------------------------------
 2 \\
 3 Paper: archive/papernum
 4 From: full name <username@address>
 5 Date: Mon, 11 Nov 2002 12:45:54 GMT (238k)
 6
 7 Title: Title
 8 Authors: Firstname Lastname
 9 Comments: 9 pages, 7 figures
10 \\
11  abstract
12 \\
13
14 Contains:
15  paperfile.tex: 104076 bytes
16  figure-1.ps: 4257 bytes
17  figure-2.ps: 10405 bytes
18  figure-3.ps: 28765 bytes
19  figure-4.ps: 19302 bytes
```

For example, in fragment 5:6, the eye of an operator will catch a misfit between the submitter

naming `"7 figures"` in the `Comments' on line 9 and a list of four files with a `".ps"` extension in the `Contains' section (lines 16-19). Seeing this misfit, the number 7 against four `.ps` files on a list, the operator is detecting the possibility of an incomplete preprint. This calls for *further inspection*, the routine method of which was to retrieve the actual preprint.

In this example, the anomaly that indicates the possibility of an incomplete submission, has not been detected by the application software. There is no specific `warning' to be seen, only something glimpsed in a mixture of auto-diagnostic data and metadata supplied by the submitter. There were a number of such scenarios possible in relation to the detection of incomplete preprints. In the absence of an explicit warning, detecting the anomaly would hinge on the ability of the operator to see where to look on the display and see what counts as potentially disorderly. Indeed, `seeing' and `looking' is what I refer to when I am talking through the inspection work as an operator (cf. Chapter 4). The following is an example of talking through a submission replacement report.

Fragment 5:7

```
1  I see a replacement size decreasing down to one tenth the original
2  submission…  (.)  comments say it is only four pages and no figures but
3  I am still going to look at its web appearance. It seems to be a whole
4  paper, nothing is missing but… this is only something the eye can catch
5  and is not flagged by the system at all.  […]
```

The news reader will thread a report of replacement with reports of previous attempts on the day, the thread being shown with an indent.

Fragment 5:8

```
1  [ 87: e-print archive       ] archive/papernum (Thu)
2        < 86: e-print archive       >
3        < 85: e-print archive       >
```

The operator can flip between the lines with the `up' and `down' keyboard arrows, and press enter to open each of them for inspection. The reference to `size' and `comments' on lines 1-2 in fragment 5:7 corresponds with lines 5 and 9 in fragment 5:6. The anomaly here is the size of the compressed packages reported by the system with the date-time stamp. The number,

shown in parenthesis, has shrunk by a factor of ten between attempts. Furthermore, the operator detects that no figures are reported by the submitter on the `Comments' line in the replaced version and so decides to inspect further, “look at its web appearance”.

As the operator reports in fragment 5:7 (line 4), “this is only something the eye can catch”. *This*, refers uniquely to a significant difference between compressed sizes of submissions of the *same* preprint, the measures of which are displayed in parenthesis towards the top of the report of each attempt as so and so many numbers of kilobytes (k). That discrepancy, “is not flagged by the system”. Also, *that*, which was not flagged by the system in fragment 5.6, refers uniquely to “7 figures“ and a list of four files with a “eps“ extension, a discrepancy which is not auto-detected either.

There were other scenarios which could not be reliably delegated to the application software. For example, the `Contains' section was routinely inspected by the operators for illegitimate file formats and combinations. A submitter might include PDF figures (figure.pdf) with a TeX source package, in some cases one PDF file with all the figures was uploaded with a TeX file. A submitter might upload one PDF version and another PS version of their preprint, neither of which had been auto-detected as TeX export. But the operators were also inspecting for single-file PDF or PS submissions which were always checked for how they were produced and if they were copyrighted, as I will come to in the next section. These particular submissions were flagged for checking in a separate overview report later in the day.

Two warnings of consequence were specifically flagged in these submissionn reports, big files and `bad' postscript output. If single figures were `looking big' they were reported “(looks big)”, although the package was still within the total size limit for processing.

Fragment 5:9

```
1 Contains:
2  figure-1.eps: 7407 bytes
3  figure-2.eps: 9425 bytes
4  paperfile.tex: 77366 bytes
5  figure-3.eps: 13784 bytes
6  figure-4.eps: 513784 bytes (looks big)
```

```
7   classfile.cls: 67724
8   stylefile-1.sty: 26264
9   stylefile-2.sty: 66574
```

This was one of the most common examples of submitters not following guidelines. It was manifested in figures that, on inspection, could be diagnosed as inefficiently produced. The instructions on the help pages suggested that 50 kilobytes of figure data is achievable in most cases. But, since submitters could be expected not to follow the guidelines, the application software was configured to report figures over the 100 kilobyte mark as `"(looks big)"`. Such flagging of figures called for still further inspection.

I will enquire into communication and corrective actions relating to big figures in the next two chapters. The other explicit warning, about a bad `postscript` output, was flagged **`"*** PS BAD ***"`** across the top of these reports if the output failed the integrity check. This would be a serious case of a preprint that, in all likelihood, would not render without error warnings in a `postscript` viewer and, if that was the case, a `PDF` conversion of it would certainly fail. Again, the routine measure in these cases was to retrieve the actual preprint to inspect and diagnose its data contents.

The examples I have now consulted on what is looked at and looked for tell us that when anomalies were detected *further inspection* was needed. A call for the actual preprint often became the next thing to do. In other words, these data reports operated as entry points to those preprint packages that needed to be looked at for one or another reason and the bridge between a report and a preprint was the preprint identifier, **`"archive/papernum"`**, shown on line 3 in fragments 5:5 and 5:6. It could be `grabbed', ie., copied, and then pasted into any available command that concerned the handling of the package. It was also encoded in the news reader as a `live' link, ie., `http://`**`archive.domain/archive/papernum`** so it could be pressed as a hyperlink, invoking a web browser. Often the quickest way for the operator was to inspect the web appearance of a preprint—to flip quickly through the pages, as the operator reports in fragment 5:7 (line 3) `"look at its web appearance"`. For example, if the

preprint was inspected for missing figures, the operator would be looking for references to figures in the text, say, `[fig. 5]' when no figure 5 was displayed anywhere. We shall now take a look at the inspection of preprints.

Inspecting preprints

In this section, I will take two examples of inspection involving a `"*** PS BAD ***"` warning and single-file `PDF` or `PS` submissions. I will begin with the former.

When preprints were submitted to the archive, a failed integrity check was often caused by embedded `postscript` figures. Sometimes these figures were corrupted in which case they had to be inspected, diagnosed and fixed. Sometimes they were so-called `encapsulated postscripts` where the core of the image data was in binary format, ie., in `GIF` or `JPEG` format. These figures were usually not corrupted, as such, but elements in the binary data were misinterpreted by `TeX` compilers. If this was the case, the final `postscript` output would be corrupted, and the package had to be recompiled with access to a special directive to ignore these binary data elements. But there were also other reasons for a `"*** PS BAD ***"` warning which could be distinguished quickly by testing if a `PDF` conversion of the `postscript` worked on the archive website. I quote myself explaining this warning while doing this inspection work.

Fragment 5:10

```
 1  Ideally with this type of warning, first [task] is to grab the paper
 2  identifier and see what it [the paper] looks like on the archive web
 3  [site]. First see if it generates a PDF [from the postscript file]. If
 4  it doesn't something is obviously wrong. It does.  (.)  …hmm
 5  interesting  (.)  …hmm…  (.)  …ok, the last page is an image, a model,
 6  it is a big image file so the ghostscript generation [test] times out
 7  and flags this warning [PS BAD]. Postscript generation is [working]
 8  alright, it simply takes ghostscript quite a while to draw that image.
 9  PDF is much faster.  (.)  And now I am going to make sure that the
10  postscript actually works… [look at it in a postscript viewer]  (.)  it
11  does.  (.)
```

First, the `ideal' next thing to do is reported, ie., to `"grab the paper identifier […] and see if it generates a PDF [from the postscript file]"` (lines 1-3). This will quickly

distinguish whether or not it is necessary to run a diagnostic device on the figures in this package to see if they are corrupted or corrupting the output. The next thing I report is that a PDF conversion has worked, meaning, that I was actually looking at it and inspecting its pages through to the last page, mumbling ″ (.) …hmm interesting (.) …hmm… (.) ″ (lines 4-5), until reaching the last page and suddenly detecting the problem: "…ok, the last page is an image, a model, it is a big image file" (lines 5-6). In lines 6-8, I explain why this is happening: "so the ghostscript generation [check] times out and flags this warning [PS BAD]. Postscript generation is [working] alright, it simply takes ghostscript quite a while to draw that image", and then I check "to make sure that the postscript actually works… […and] it does" (lines 9-11). At that point, I had an uncorrupted preprint open in a viewer on the display and was actually looking right at it.

To sum up, the possibility that a postscript might be corrupted was manifested in a "***** PS BAD *****" warning. But evidence that the postscript was indeed an illegitimate object was quickly verified if PDF conversion on the archive website failed. This would be manifested in the fact that no PDF opened in a relevant viewer and the operator or the submitter would see a message in a web browser announcing that `**PDF is unavailable…**' This was a preferred choice, to first verify whether or not there was a corrupted code somewhere. It was also quicker than flipping through the pages of the postscript just to see whether or not a window with error code would pop up on one of the pages, indicating the location of `bad' data. But, if bad data was actually the case, and the submitter had failed to correct the problem as the 4pm freeze was approaching, the operators had two options. They could remove the package from the pending queue or copy it from the main server to inspect, diagnose and fix its contents.

The inspection of single PDF or PS preprints presented a very different set of potential disorders. Copyright issues are non-negotiable for legal reasons but the evidence of copyright is not reliably machine-readable nor are all TeX produced exports reliably detected by

software. Therefore the operators manually inspected all single-file PDF or PS submissions. Submitters might have uploaded a journal-generated PDF copyright of that journal; they might have scanned an article, a journal copy or not, and saved it as a PDF or PS file Submitters might have used postscript (T3) fonts (bitmap fonts) in preparing the preprint in some word processor or TeX typesetting software from where a PDF or PS file was exported. These files in particular could be illegible at certain resolutions.

If a PDF or PS upload was not auto-diagnosed as TeX produced and consequently auto-rejected, detecting and inspecting these files called for combined efforts. The apparatus detected a single PDF or PS upload and reported their presence in the system, while an operator responded to these warnings by inspecting the files in dedicated viewers. Single PDF files were far more common than single postscripts. I quote remarks by an operators who was responding to my question about the presence of postscript (T3) fonts in a single PDF preprint where no visible evidence of a `bad' font was detected on manual inspection.

Fragment 5:11

```
 1 unfortunately there is no hard and fast rule. If there was, we could
 2 autodetect and autoreject and save us a lot of time. A simple
 3 occurrence of a T3 font does not necessarily mean that the PDF needs to
 4 be rejected.
 5
 6 The manual pdf inspection has several objectives:
 7
 8 detect badly generated, i.e. mostly illegible, pdf of any origin
 9 detect tex generated pdf
10 detect third party copyrighted, e.g. journal generated pdf
11 […]
12
13 archive work will always involve judgment calls.
14 […]
```

The operator claims that “there is no hard and fast rule” (line 1), and uses the example that occurrences of T3 fonts are not clear-cut for rejection (lines 2-4). The operator also lists the objectives associated with “manual pdf inspection” on lines 8-10. The named objectives are to detect “illegible” files, although they could be “of any origin”, “tex generated files, and “third party copyrighted” files. The final remark is perhaps stating the obvious, that this “work will always involve judgment calls” (line 13).

Listing these objectives does not specify what an operator would be looking at other than the example, “illegible”. To avoid borderline judgements about legibility, the rule of thumb was to remove all scans and all files with T3 bitmap fonts in the main text. The former would be detected by looking at the pages of these files to see if they showed, for instance, dark edges, slight blurring of the text, no scalability if zoomed in and out, or if the formatting of title, abstract, etc., indicated a publication. These phenomena would immediately give the appearance of a scan from an existing hard copy. The absence of metadata, listing fonts used in the generation of the file, would also give it away. Both viewers and a command-line argument can extract that data if it is present in the file. The use of T3 bitmap fonts can be quickly detected by zooming in and out of the main text. Bitmap fonts do not scale. But metadata on the use of these fonts is also present in the file's metadata although they might refer to text in figure captions or other text which is not the main text.

The same or similar phenomena were detected when single PS files were inspected. There were no hard and fast rules as the operator remarked. Rather, establishing evidence that single-file PS or PDF preprints were not according to guidelines had to be found and seen in the details of their pages

Inspecting the message box

As we have now learned, inspecting the metadata reports and the preprints was one way to curb problems that were not targeted by the submitters—to inspect for machine-reported warnings and any other disorderliness which had not been auto-detected. But the metadata reports also served as a resource for *having a sense* of the goings-on during the day. They gave a sense of data traffic, for instance, how many packages were coming in and from where. Repetitions could easily be detected such as if the same oversized photograph of some phenomenon was circulating amongst astrophysicists, ie., same file name, extension and size in the `Contains' sections, or if there were suddenly many instances of a TeX auxiliary that

was causing trouble. Records also show that a day with 10-15 auto-detected single PDF uploads awaiting manual inspection was a number in the higher range so it was helpful to notice these uploads early in the day, rather then wait with the checks until they were reported later. For example, one day on record shows over 40 single PDFs coming from participants at a conference which was respectively mentioned in the `Comments' sections of the bibliographic data. These conference participants were computer scientists not using TeX.

The principal actor made choices about what to tackle first and what to leave for later in the day, even what to leave until after the 4pm freeze. It was not the sort of work one could organise in any strict sense. Rather, the principal actor was a *triage* person, someone who begins by inspecting the situation, then prioritising the work that needs doing, but the priorities would change over the hours. In other words, this work was sensitive to both the time constraints of the day and demands that emerged as the time passed. A good part of this work was also to inspect the message box for indicators of what needed attention or might take place later in the day. The box was inspected throughout the day for messages from submitters with questions, complaints or requests, or messages from other operators, moderators and miscellaneous others.

For example, on one busy Monday, a few communication threads were awaiting attention, some dating back to the 4 pm freeze on the previous Friday. For an operator who had been away or not been lurking in the background for a few days, coming back on site always required a quick skim through the message box. It was critical at least to look at messages and communication since the last freeze in order to have a sense of events that might have impact on scenarios during the day. I will give an example. **AA** is the principal actor, **AB** and **AC** lurk in the background, S**01** is a submitter, and **SYS** is the system.

Fragment 5:12

```
1  16:57  S01: to admins
2
3         […] I would like to submit my Phd thesis to the archive, but due to
4         its size… […]. It contains many figures (more than hundred) that I
5         have already shrunk by converting them into gif files and then to
6         compressed postscripts. […]
```

```
 7
 8
 9 21:19  SYS: to admins
10
11       S01@university fpu has been piped through manually to archive
12
13
14 21:26  AA: to S01 (copy to admins).
15
16       Now archive/papernum
17       Removed [the figures] figure1.eps figure2.eps figure3.ps which were
18       not used
19       Removed [the temporary processing file] file.flg which must not be
20       present
21
22       Note that we do not run metafont here, if this has caused problems you
23       must include the appropriate output files from metafont.
24
25       You should have received a separate email notification of acceptance
26       which includes the paper password. Please use this to check the
27       submission.
28
29
30 21:27  AA: to admins (subject: RE: AB/AC -> […]).
31
32       first problem was reported in [quoting] log: can't write file.flg [end
33       quote] which was due to it [the file] already being in the submission.
34       other files needed ignore setting [a directive inserted for the auto-
35       compiler to ignore certain elements in binary image data]. may still
36       have font probs but I'll leave that for the author to check.
```

After last Friday's freeze, **S01** reported an oversized package that failed to upload (lines 3-6). The submitter explains, "[i]t contains many figures (more than hundred)" (lines 4-5) which have been compressed, the details of which correspond with instructions on the archive website, ie., by "converting them into gif files and then to […] postscripts" (lines 5-6). A record of this package is visible again when **SYS** reports it being "piped through manually" (line 11). **AA**, who has inspected the contents, reports to the submitter a few minutes later. There is a mention of removal of three superfluous figures but there were other problems. What **AA** did and what still might need doing is explained to S**01** in this letter which is copied into the message box (lines 16-27). Do not include **file**.flg and upload your own output from metafont. A minute later **AB** also reports directly to the administration on what is the case, including directives with reference to the use of `**file**.flg', `ignore settings' and potential `font probs', in case the author needs to resubmit (lines 3236).

The problem is taken care of on that Friday night but making a note of the exchange was important in the event of seeing further attempts by S**01,** who now has until 4 pm on Monday

to check the outcome and resubmit with `"the appropriate output files from metafont"` if necessary. If that was to happen, the new submission would still be auto-rejected due to its size and again come under scrutiny of an operator. But the operators now had a recorded history of communication and interactions accounting for the manoeuvring of this particular package, including the auto-processing logs which were omitted from this exchange but readily available. In other words, the entire record is an available future resource.

For **AA** and the other operators, this particular communication does not convey any previously unknown technical facts about `TeX` processing on the archive or workarounds with `` `metafont' ``. It is simply a reminder, one of many that would be found in the message box, of what to look out for and be sensitive to on the day. This will become much clearer in the next chapter on *taking corrective action* but, insofar as the message box is inspected in order to detect anything amiss, there are two auto-generated overview reports to consider. Anomalies and diagnostic data on activities during the day were harvested in the last hour before 4 pm every weekday. They were gathered into a couple of reports and routed into the message box, first 45 minutes and then again 5 minutes before the 4pm freeze. Some items are listed for auto-correction and others for manual inspection. This way, potentially problematic items, if they are not taken care of already, are `shovelled' into immediate focus for a second round of inspection. I will display a small excerpt.

Fragment 5:13

```
 1 Here is today's e-print report:
 2
 3 Automatic Errors and Warnings
 4 RUN /tmp/autofixes to fix ... list of archive identifiers
 5
 6 Manual checks:
 7 astro-ph/papernum: possible 3rd party submission from
 8  > full name <name@address>
 9  > Authors: author name ?
10 astro-ph/papernum: HTML accent: &uacute; ?
11 ^^^^^^^^^^^^^^^^: HTML accent: &iacute; ?
12 ^^^^^^^^^^^^^^^^: HTML accent: &aacute; ?
13 astro-ph/papernum: contains exclamation mark !: ...quoting the abstract ?
14 astro-ph/papernum: author with only surname? ...a list of author names and
15 institutions ?
16 astro-ph/papernum: PS only submission, check it ?
17 […]
18 […]
19 cond-mat/papernum: SINGLE PDF file, check it!
```

```
20 cond-mat/papernum: seems short, check it:  2 pages
21  >Title: [title] ?
22 […]
23 hep-ph/papernum: unnecessary formatting in metadata: .~\ ?
24 ^^^^^^^^^^^^^^: : cite \cite{ref.3} is not human readable?
25 math/papernum: Withdraw notice, check it ?
26 math/papernum: Title starts with a lower case letter: p-Modular ...quoting
27 the title ?
28 […]
29 Cross Listings (only papers with >=3 crosses)
30 […]
```

Lines 3-4 in a section called `"Automatic Errors and Warnings"`, remind the operators to `"RUN [the script …] autofixes to fix"` a batch of bibliographic data files. This is often a long list of items which mostly have obvious typographical errors, extra white spaces, and excessive use of `TeX` markup of formulae and citations in abstracts. But `` `autofixes' `` cannot reliably recognise and auto-correct errors, thus when an operator invoked the script on the command line it would display suggested changes to one file at a time, before and after, awaiting confirmation: yes, no or manual edit.

A section called `"Manual checks"` (lines 7-70) displays a list of uncertainties which I have shortened significantly in this fragment. Some of these are uncertainties about bibliographic data which have not, or cannot, be delegated to `` `autofixes' `` and some of them regard data packages that require manual inspection. A registered member is detected as 3rd party to the author list (line 7). An `html` encoding of letters with accents: ú, í and á, is found (lines 10-12). An abstract `"contains exclamation mark !"` (line 13) and line 14 indicates that something is amiss with the author list `"author with only surname?"`. Then there are `"PS only [single postscript] submission[s]"` and `"SINGLE PDF file[s]"`, which are always reported (lines 16 and 19). A document `"seems short"` (line 20). Lines 23 and 24 concern `TeX` markup in abstracts, here the markup of hyperlinked citations, `"cite \cite{ref.3}"`, the system detects a `"[w]ithdraw notice"` (line 25) and line 26 displays the opening of a title with the technical term `"p-Modular"`. These have to be approved or edited. Finally, line 29 heads a section containing all cross listings between the subject archives on the day. They are inspected in case they are excessively used or not topic-relevant.

These auto-diagnostic reports were diligently inspected by the principal actor of the day who would run `autofixes' and manually check off each of the items under the heading, "Manual checks". All items on these lists would actually call for further inspection. Either the actual preprint files had to be inspected or the bibliographic files, **papernum**.abs, if necessary, were manually edited with a command-line editor on the main server via terminal access.

I will come back to further analysis of these particular inspection/detection activities in the final section of this chapter but, so far, we have learned that inspection and detection is solitary work. This is true for the most part because each of the operators were physically by themselves. It was also the case, as I have explained, that one of them was the principal actor on the day but with one or two operators lurking in the system and keeping track of communication about problems with submissions. Operators would find themselves examining the message threads intermittently during most days. And, there was always some communication between them about the daily administration of preprints, especially in relation to uncertainties. The next section will take examples of inspecting together.

Inspecting together

In this section, I will examine two examples of communication involving assistant `techies' who were still learning what to look for. These are examples of simple oversight which were not uncommon in using the computer supports. They also provide particular visibility of where to look and what to look for.

AS is the assistant techy, reading through the metadata reports, and **AA** is an operator on site on the day.

Fragment 5:14

```
09:20     AS:  to admins (subject: PS BAD, but the paper is fine.
            (archive/papernum)).

            […]
            This paper contains ps bad. (archive/papernum).
            However, I checked through both PDF and PS file and they are fine.

            Should I ignore the PS bad warning in this paper then?
```

```
 9            […]
10
11  13:58     AA:  to admins (subject: Re: PS BAD, but the paper is fine.
12            (archive/papernum)).
13
14            The submitter fixed the problem in the second upload
15
16              [ 103: e-print arXiv       ] archive/papernum (Mon)
17                   < 102: e-print arXiv       >
```

AS had come across a **"*** PS BAD ***"** warning in one of the reports, as indicated in the subject lines (1 and 11). **AS** had not found anything amiss with the preprint and now asks if the warning should be ignored. One of the operators responds a few hours later in the day by quoting directly the index for threaded reports which shows a report for this package and its replacement (lines 16-17). On inspection, the latter attempt did not fail this routine integrity check, ie., no **"*** PS BAD ***"** warning. **AS** could have seen this, but had not done so at the time. The direct quote from the records index of incoming submissions however, directs **AS**'s attention to evidence of the submission's history, ie., replacements indicated with the indentation in the list providing access to each upload report independently.

It should be mentioned here that the operators were virtually interchangeable in the daily administration of preprints when I first came on board. Divisions emerged however, with respect to experience and technical expertise when assistant techies came on board. This is not to overlook that the operators were already differently positioned in relation to one another in the sense that some divisions had emerged and been established through their shared histories of solving problems on the archive. Someone was always lurking in the background and available as in the example above.

Another thing we learn here is the practical significance of using the submission identifiers to communicate, **"archive/papernum"**, in the subject lines (fragm. 5:14, lines 2 and 12; fragm. 5:15, lines 2 and 8). The identifiers were inserted into the subject line in communication relating to specific preprints. In the same way as the identifiers served as a bridge between diagnostic reports and the actual preprint packages, they served in the communication as a bridge between individual messages and the preprint package the

communication was about. This arrangement provided direct access to the preprints that were being scrutinised as well as providing immediate overview of communication threads about particular preprints, since the subject line would always be visible in the message index on the day.

The second example I will take in this section poses a somewhat different set of questions, more specifically, the question of how computational devices did their share of the work. The example I will take concerns two of the most frequently used home-made tools `hack' and `pscheck'. More specifically, `hack' was script used to copy submissions over from the main server and uncompress it. The other script, `pscheck', was used to inspect the syntax of postscript figures embedded in submission packages when the TeX auto-compiler reported **"*** PS BAD ***"**. Three operators are **AS**, the assistant techy, **AA** and **AB**.

Fragment 5:15

```
 1  13:38     AS:  to admins (subject: hack/pscheck (archive/papernum)).
 2
 3            When I run [… the script] pscheck on this, I get :
 4            ___________
 5            -rw-r--r--  1 [uid]  [gid]     847 Dec 10 13:25 papernum.abs
 6            -rw-r--r--  1 [uid]  [gid]   48482 Dec 10 13:25 papernum.tar.gz
 7            -rw-r--r--  1 [uid]  [gid]   67724 Mar 26  2001 cjour.cls
 8            -rw-r--r--  1 [uid]  [gid]   26264 Dec  8 00:53 cjourps.sty
 9            -rw-r--r--  1 [uid]  [gid]   66574 Dec  8 00:53 cjour.sty
10            -rw-r--r--  1 [uid]  [gid]   24505 Dec  8 00:53 paper.tex
11            AS@terminal:/home/AS% pscheck
12            > foreach: No match.
13
14            ___________
15
16            Document views ok otherwise. […]
17
18
19  13:42     AA:  to admins (subject: Re: hack/pscheck (archive/papernum)).
20
21            If you look at [read] pscheck you'll see that it does:
22
23            foreach f (*.ps *.eps *.PS *.EPS *.epsi *.EPSI)
24
25            it only attempts to look at PS files and this submission doesn't
26            contain any.
27
28            […]
29
30            Looks okay to me. What were you attempting to check with pscheck,
31            was there some page of documentation that mislead you?
32
33
34  13:45     AB:  to admins (subject: Re: hack/pscheck (archive/papernum)).
35
36            pscheck does a syntax check on postscript files in the current
```

```
37          working directory. I don't see any postscript files in the above
38          list, hence the error message is correct.
```

The exchange ends here. **AS**'s message cites a report from the script `` `hack' `` and then from `` `pscheck' ``. A list of files, extracted from **papernum**`.tar.gz` by `` `hack' ``, is on **AS**'s display (lines 5-10). They belong to a submission package that **AS** has copied or *hacked* over to a local desktop for inspection and diagnosis. Then **AS** invokes `` `pscheck' `` (line 11) which reports back that it finds `"[n]o match"` (line 12). In response to this message, **AA** invites **AS** to read `` `pscheck' ``, ie., `"[i]f you look at pscheck you'll see [… what] it does"`, citing a section of the script (line 23). A loop `f' is defined and `for-each' item to be routed into this loop, parameters are given as to how the file format is identified. Certain file extensions in this case, `"(*.ps *.eps *.PS *.EPS *.epsi *.EPSI)"`, are the relevant cues for `` `pscheck' `` to include a file in the loop and embark on syntax inspection. **AA** points out to **AS** that `"this submission doesn't contain any" postscript` files. They agree that the document views fine (lines 16 and 30), but **AA** wants to know what led **AS** to use the script in the first place (lines 30-31). This question was never answered.

A closer look at this exchange illustrates that operators are expected to look for and see clues that can clarify a problem. Rather than simply stating the solution, **AA** demonstrates this craft. In his response, a section of the code of `pscheck` is cited (line 23) to allow **AS** to `"look at [read]"` the script and see what it does. This deliberately brings into focus a comparison with **AS**'s earlier reference to the file list reported by `hack`. On the list of files that were made available to `pscheck` (lines 5-10) none have an extension that matches the parameters named on line 23. In lines 36-38, **AB** gives an explanation of the role of `pscheck`, `"[P]scheck does a syntax check on postscript files in the current working directory"`. **AB** also refers to the file list on lines 5-10 as sitting in the `working' directory, `"I don't see any postscript files in the above list"`, and concludes that `"the error message [produced by the script] is correct"` (lines 37-38). In other words, there is nothing wrong with `` `pscheck' ``.

This naming and pointing at observables, whereby operators attempt to see what is going on, manifests their intimate knowledge of and confidence in the devices they use. Most obvious here is the citing of a section from the code and the suggestion to `look' at it and then `you'll see'. There is repeated referencing to what can be observed about `pscheck', what it does in this case and why it reacts as it does. The possibility of `pscheck', or one of its functions more precisely, acting somehow *not* according to design and configuration is systematically and collaboratively eliminated between **AA** and **AB**. It is made clear as well what **AS** fails to see on the display. What is not visible, however, is what led **AS** to invoke a circumscribed `postscript` syntax inspection when none of the files on display are `postscripts`. The probe from **AA** about `“some page of documentation”` (line 31) is an attempt to discover what *irrelevant* information **AS** might have been mislead by. I will now turn to my discussion of what we have learned about the work of inspection and detection and relate this work specifically to matters of practical responsibility and guidelines.

Discussion

In the first sections of the chapter, we looked at two auto-responses explaining to submitters what to do and what to expect when preparing and submitting their preprints to the archive (fragment 5:1, 5:2 and 5:3). As we saw, these responses describe how to be a responsible submitter, the details of which are to pay attention, check, verify a host of named items and take corrective action. For example, the list of items in fragment 5:2 is detailed in naming those things that should be correct, complete, in order, standard, classified and uncorrupted. What I have drawn attention to in the main sections of this chapter however, is that which, on inspection, is incorrect, incomplete, corrupted or irresponsible.

The incentives submitters should have to make sure their metadata and preprint files were in order did not always materialise. Neither did the explicit warnings stipulated in the auto-responses in fragment 5:1, 5:2 and 5:3 have the effect of eliminating erroneous

submission packages. Achieving a seamless mass-dissemination of preprints in the service of thousands of scientists hinged on inspection and interception on the back end in response to the fact that submitters were sometimes unreliable.

The `ideal' or `orderly' submitter, if such a person indeed exists, could only be realised by reference to preprints and metadata which barely came to the attention of operators. This orderly submitter did not introduce deviations or anomalies into the daily grind. In fact, it should be noted that, of an average of 140 submissions a day, most submissions were taken care of at the front end. But, as I have explained, my central concerns are problems of disorder which, in this chapter, attach themselves specifically to the anomalies we now know operators and their computer supports are detecting on the back end.

To contrast orderly submitters / submissions with disorderly ones invites a familiar mode of sociological reasoning. For example, why are these goings-on not contextualised in a frame of normative reasoning about practices of writing scientific articles to which registered submitters normally adhere while a small number deviate? In Chapter 2, I discussed concepts and arguments explained by the founding fathers of sociology, and looked at some of the developments within sociology that draw on them. In Weber's account of sociology, subjective meanings are attached to social actions which are normatively oriented in their course to pre-existing social structures, circumstances and constraints. Within this framework, deviation and disorder can only be defined and evaluated on the basis of *unusual* circumstances and origins, ie., in terms of pathology.

We could argue that in the pre-written responses to submitters, the operators construct and impose not only technical guidelines but also normative and moral guidelines. They attempt to commit submitters to obligations to orderly conduct and constrain them to vigilance and other responsibilities. The fact that submitters cannot be relied upon results in these responsibilities being imposed on the back end instead. What counts as reliable mass-dissemination of scientific articles to thousands of scientists is to provide uncorrupted and appropriate preprints *en masse* on a continuing basis. These are significant constraints and

conditions guiding the work that needs doing on the back end but the actions of operators and submitters also have impact on how this preprint service evolves.

The problem with this line of argumentation is twofold: the first lies with method and the second with particular ontological assumptions as I discussed in Chapters 2 and 4. In this case, the researcher will, in all likelihood, be faced with tensions between the meanings submitters and operators attach to submitting and preparing preprints. They would be facing the classic tensions of the system-actor, agent-structure dichotomies and there is enough data available to tell a convincing story addressing common topics of moral responsibility, power relations and more. Such a mode of reasoning however relies on formal structures of internal-external relations, norms and deviance, socio-historical conditions affecting the course of local activities and local activities affecting the course of history. But to draw such ties between circumstances and what is happening at the local level needs to arrive at causal explanations and adequate evidence which is only achievable by distancing the observer and creating a *standpoint*. From this standpoint persons, things, motives, meanings, constraints, circumstance, power dynamics, morals and disciplines can be defined and characterised, and their relations explained. I addressed throughout Chapters 2-4, my objection to this mode of method and reasoning, ie., how the study policies of conventional sociology draw attention away from the detailed phenomena of the actions and interactions of individuals going about their business, as phenomena to be studied in the their own right to find in them richness and specificity with regard to the ties between circumstance and those lived details.

What I did in the main sections of this chapter was to layout the work of inspection and detection. What I argue is that we do not need conventional sociology – method and reasoning – to uncover the circumstances of doing this work. The work we can see the operators and their computer supports do is furnished with details that are specific to questions, to paraphrase Lynch (1982), of the contingent ties between circumstance, origin, constraints and more, and what is flagged, seen or otherwise detected as *not in order*. For example, the operators commit to questions of completeness and origin of the objects they

inspect. In the section on `inspecting metadata reports' we found operators attending to the details of a mixture of auto-diagnostic data and metadata supplied by the submitter. A question of completeness attaches itself to what can be seen, for example, the discrepancy between the number of kilobytes reported between original submission and replacement (fragment 5:7). It is in seeing these `things' – *there* on the screen in *that* report – that other phenomena in the same report are circumstantially relevant. The archive identifier, **`archive/papernum`**, routinely reported on line 3 in these reports (see fragm.5:5), is the link and access to the preprint's origin in the system, from where it can be retrieved to check whether or not it is in fact incomplete and in *what* the origin of the anomaly lies. It is also in seeing and detecting these details that the tie to the preprint's originator, routinely reported on line 4 in these reports (see fragm.5:5), is circumstantially relevant in case the submission has to be removed. In other words, these ties are contingent and specific only to the practical relevance of origin and originator *there* and *then*.

In doing the inspection work, the operators also commit to questions of constraint which is material as well as formulations of what is or is not according to guidelines. Is the preprint corrupted or badly generated? For example, in the section on inspecting preprints, we learned that concrete evidence of a preprint being scanned from a hard copy or containing bitmap fonts in the main text, are `slight blurring of text' or `no scalability if zoomed in and out'. There were specific requirements published of what a preprint had to conform to— requirements formulating, among other things, that preprints should be legible, using scalable fonts and not be copyrighted to journals. But these particular requirements are distinctly formulated in the very act of consulting documentary evidence that can be seen on the pages of preprints, judgements made *there* and *then* that they are not in order. On the basis of *that* detection work and judgement, they would be removed and the rejection explained to the submitter.

The evidence of a scan or bitmap font are manifestations of particular material constraints of producing `PDF` or `PS` files from a scan or with fonts that cannot scale. These

constraints present themselves in a preprint file that cannot hide its origin and the source of its appearance. The explicit warning, `*** PS BAD ***`, is also a manifestation of material constraints in the sense that a `postscript` output cannot pass an automated integrity check. But, as we learned from fragment 5:10, the origin and source of that warning is hidden. We also learned that the operator works the constraints that are known about the application software, `"see what it looks like on the archive website [...] see if it generates a PDF"` (fragm.5:10, lines 2-3). In doing this, the operator is inspecting for the origin and source of this warning and can verify whether or not a corrupted code is present in the `postscript`, thus whether or not further inspection and corrective action is needed. What we see in these examples comes back to the matter of practical responsibility as one that attaches itself to what the metadata and preprint files are inspected for, what is detected in them and if they are removed. The matter is raised with submitters in the auto-responses we saw and it is also raised distinctly with respect to phenomena manifesting what has been ignored or overlooked by submitters.

We have also seen that in order to keep track of incoming submissions, inspect them and detect for anomalies, involves computing support. We learned that there are no hard and fast rules for finding that which, on inspection, can be determined as a disorderly object. As the operator puts in fragment 5:11, this is unfortunate. The processing of metadata and preprint files, that are reliable and consistent in the strictest sense, was automated. But there are obvious limitations to the support that the system provides. For example, certain things could not be flagged at all as we saw in fragment 5:6 and 5:7 where only the eye can catch something amiss. In other cases, what is formulated as potentially incomplete is reported by detecting specific material attributes of preprint files. For example, `"2 pages"` is reported with a canned message `"seems short, check it"` (fragm. 5:13, line 20). That object could potentially be an advertisement or errata and not a scientific article, although they are sometimes only 2 pages. Another example is `"contains exclamation mark !"`, followed by a

brief quote from the abstract (fragm.5:13, line 20) and potentially indicates that an author is using inappropriate statement or somehow drawing attention to themselves, neither of which is good publication practice among physicists and mathematicians. But exclamation marks are also routinely used as operators in formulas. In this we see how deviation from convention was formulated in computable rules for detection of anomaly, rules which could only be invented ad hoc. In the former example, operators had learned that notices and errata were being submitted. In the latter example, they had learned that submitters may occasionally make inappropriate statements using exclamation marks. By flagging all objects 2 pages or less, and all abstracts with exclamation marks, they had invented `anchors' for checking culturally-specific formulations *there* and *then*. Is this a preprint? Should a moderator look at this preprint?

The operators relied on communication protocols to route the entire data traffic through filters of auto-diagnosis, reporting and manual inspection, work which otherwise was done in isolation. But they also relied on communication protocols to mediate any communication about procedures when needed, and keep track of histories and prospective futures as we saw in fragment 5:12. Furthermore, we learn in the previous section how operators rely on intimate knowledge of what the devices they use are actually doing (fragm. 5:14 and 5:15). The examples I took involved a member on the team who was seeking assistance. We see how they rely on each other to name and point at observables manifesting, for example, a replaced submission in fragment 5:14 (look for the indentation) or in fragment 5:15 when elaborate efforts are made to show what the script `pscheck' is awaiting to detect, and what file types are made available to it.

The aim of this chapter was to specify problems of disorder by explaining and allowing the reader to see what operators were looking at and seeing. But in specifying these problems the analysis has also shown some of what the computer supports are configured to diagnose and detect. What we see is a complex relationship between computational functions and subtle human judgements which together accomplish this work on an on-going basis, and

with reference to issues of practical responsibility and adhering to guidelines. This relationship would be missed in normative reasoning about social action, the order of a practice and technical work. Also, we can see that procedures and the use of computer supports are implemented ad hoc to accomplish this work. In the next chapter, I will continue this line of reasoning and argue that work procedures and the use of supports have no place in formal descriptions, task analysis or assumptions about cognitive control structures in the order of work.

Chapter 6

Taking corrective action: trial and error

Introduction

As we learned in the previous chapter, procedures which were initiated by submitters on the front end draw attention to the work that needed doing on the back end. The bibliographic data were inspected. The preprint files were inspected. Machine auto-diagnosis and reporting was inspected. But, as we will now see, inspection and detection does not stop here. Inspecting for and detecting anomalies is not only integral to decision-making on whether corrective action needs to be taken. It is integral to *taking corrective action*.

In this chapter, I will enquire into procedures of taking corrective action and of communicating problems and their resolutions. I will focus in the main on common scenarios that required direct communication between operators and submitters, and direct involvement on part of the operators. One of the things I will focus on are the resources to do the work of preparing preprints for submission. Many of these resources were referenced in guidelines, help texts and direct communication about particular problems that would occur once the submissions were uploaded. Messages between submitters and operators contained questions, suggestions and explanations, and included processing logs and links to help texts. I will enquire specifically into *making efforts* and *competence* as significant and socially relevant matters. They are relevant in showing how operators and submitters cooperate and cope with the archive apparatus, parts of which were routinely invoked to do their share of the work.

Your best efforts

In this section, I will further discuss the expectations communicated to newly registered submitters. I will begin by quoting again from the auto-response upon submitter registration

with the archive. The message continues:

Fragment 6:1

```
 1  It is frequently more efficient to consult a colleague first before sending
 2  email to the server admins, so please only email questions which are
 3
 4   - not explained in the online help
 5   - cannot be solved with a little trial and error
 6   - remain mysterious even after consulting with a computer savvy colleague
 7
 8  Note that on the day of submission, before the 16:00 US Eastern time (EDT/EST)
 9  deadline, you can replace as often as necessary to debug layout problems
10  interactively and to make editorial changes. There is no penalty for
11  multiple same day replacements and no new version number as long as
12  the replacements arrive here before the above daily deadline.
13
14  If despite your best efforts you cannot resolve problems with your submission,
15  send a concise description of the problem to www-admin@archive.domain, always
16  remembering to mention the archive/papernum or temporary identifier, and
17  someone here will reply, typically (but not always) within 1 working day.
18
19  DO NOT under any circumstances send your submission or any unsolicited file
20  attachments to www-admin. This is a group address only for communicating
21  e-print server related problems and suggestions. Regular submission
22  attempts are cached with a few day latency and we need only the identifier
23  you've received in order to inspect your attempted or successful submission
24  for problems.
25
26  Always contact www-admin@archive.domain if you think you have found a genuine
27  bug which can be reliably reproduced, and you have verified that your web
28  browser and display software is up to date. […]
```

The first two lines tell the newly registered submitter that consulting a colleague is `"frequently more efficient"` than emailing questions to the archive administration (lines 1-2). I draw attention to the three listed options for checking any questions that the uploading procedures might prompt, ie., `"before sending email to the server admins"`. The questions are:

```
- not explained [or answered] in the online help
- cannot be solved with a little trial and error
- remain mysterious even after consulting with a computer savvy colleague
```

Arguably, the anonymous voice of this auto-response takes certain kind of disciplined work practice for granted. Openness to experimentation and improvisation, `"a little trial and error"`, is expected of them and, as physicists and mathematicians, they belong to data and computer-intensive research professions with long standing experience in complex data management and processing. Hence, `a computer savvy colleague"` *can* be consulted. I also

learned from the other operators how they assumed that the application software on the archive should not be much of a mystery to submitters, in particular, not the `TeX` auto-compiler. Submitters should have prepared their packages by testing them on their office or laboratory computers, and they should be equipped to recognise and repair auto-processing failures on upload `“with a little trial and error”`, as the response suggests. Overall, submitters should not have to be trained or educated by the operators about the archive system and, according to this auto-response, they were explicitly told to use their problem-solving capabilities.

The remainder of the message contains a few guidelines to keep in mind when uploading packages in the future. Authors can, for example, replace new submissions `“as often as necessary to debug layout problems [… and] make editorial changes”`, and there are time constraints, or `“the 16:00 US Eastern time (EDT/EST) deadline”`, with respect to replacements of new packages (lines 8-12). The operators do recognise the possibility of a genuine problem however, and `“will reply, typically (but not always) within 1 working day”` (lines 14-17). Importantly, submitters should not (`“DO NOT”`) under any circumstance attach a problematic package in email to the admin email address because the automatically generated `“archive/papernum”` or `“temporary identifier[s]”` are used to trace packages and inspect them (lines 19-24). Finally, submitters are asked to report any `“genuine bug which can be reliably reproduced”`, and thus also contribute to ongoing scripting, realignments and reconfiguration of the apparatus (lines 26-27).

Again, it appears as if submitters are left to their own devices at the front end interface. The registration auto-response explicitly *contractualises* the new member by assigning certain tasks to submitters and potentially clearing the operators of service duties. Actually, both the registration auto-response and the submission response, quoted in the previous chapter, formulate a practice to which scientists and mathematicians of various specialities belong. The archive, as a *self-service*, had become indispensable for the ongoing exchange of

research results and many had long-standing experience with the archive as frequent visitors and submitters. The vast majority of preprints were recorded as co-authored, but submitters also hinted that they worked collaboratively and shared facilities with other colleagues. If the registered author-submitter failed, one of the others should be capable and also readily available for obvious reasons if they were co-authors and wanted to see the product published. If they failed completely, the result could be no posting on the archive.

Generally, and this is evident in the metadata reports that list replacements, submitters improvised and experimented to solve upload problems. But there were new member subscriptions all the time, the use of the system had been spreading steadily (see Chapter 1, fig:1.2), and not all of the new submitters would put in the time and effort to study in-depth how to use the application software for building and submitting preprints using `TeX`. There were some who needed assistance with their uploads. Some never used `TeX`. There were problems with oversize figures and authors communicated what they saw as problems with the archive. For instance, astrophysicists openly complained about having to compress their figure plots and colour photos *more efficiently* according to the demands communicated in guidelines and by archive operators.

Statements, demands and discouraging remarks take on the role of disclaimers. Assistance should not be expected unless `“despite your best efforts you cannot resolve problems with your submission”`. The operators frequently stripped their communication with submitters down to assessment and correction. If submitters failed to produce an uncorrupted package or if they failed to comply with file/package standards, submissions were removed from the pending queue and a brief message sent to submitters stating the reasons why. If rules were blatantly broken, copyright for instance, submitter status was temporarily suspended. But the server environment did not support completely this configuration of being a submitter. Technically compatible data components and processing capabilities are available in great numbers, and impossible to know everything there is to

know about it. As we learned in the previous chapter, the interfaces to the self-archival module were *materially* inclusive in terms of how and what could be successfully uploaded – technically speaking – a condition that invited a myriad of probable mistakes and of breaching some guideline. However, the expectations communicated to registered submitters, quoted in the auto-response in fragment 6:1, suggests that submitters *can* solve problems with their packages. They *can* look for explanations in the `"online help"` and they *can* `"consult a colleague"`. Otherwise, `"[i]f despite your best efforts you cannot resolve problems […] a concise description of the problem"` *can* be sent to the email address of `"server admins"`. Regarding the contents of some of the messages submitters where emailing, operators might remark among themselves that submitters were not paying attention, not properly consulting the help text guidelines, and not making their best efforts, but there were also issues with software applications, how they were configured, interfaced and controlled.

Resources

In this section, I will reconstruct the preparation work needed to produce a submission package in `TeX` with figures. Preparing scientific papers for publication using the scientific typesetting system `TeX`, typically requires a bit of tinkering in order to get the processing of the markup language to render properly: text, formulae, tables, figures, hyperlinks, and so on. This applies particularly to all sorts of layout concerns either as personal preference or involving layout auxiliary components from journal publishers. But this can also be an issue of getting the markup language correct so that processors can do their job without a halt. I will construct an example of a simple `TeX` source structure.

Fragment 6:2

```
1  \documentclass{aa}
2  \usepackage{graphicx}
3
4  \begin{document}
5     \title{ title }
6
7     \author{ author 1 \and author 2 }
8
```

```
 9     \institute{ institute }
10
11  \abstract{ abstract }
12
13  \maketitle
14
15  \section{ section title }
16   section text
17
18  \begin{figure}
19  \includegraphics[]{ figure.ps }
20  \caption{ figure caption }
21  \label{ label }
22  \end{figure}
23
24  \section{ section title }
25   section text
26
27  \begin{thebibliography}{}
28  \bibitem[]{}
29   reference 1
30
31  \bibitem[]{}
32   reference 2
33  \end{thebibliography}
34
35  \end{document}
```

The top line declares what document class this is, "`\documentclass{aa}`". In this case, the source will invoke a class file of the journal, Astronomy & Astrophysics, `` `aa.cls' ``. Because the class was commonly used, it was by default available to the auto-compiler on the archive and invoked automatically, unless submitters included their own copies or versions, in which case the auto-compiler would detect that and use that class instead of the local version. Line 2 invokes a standard graphics package which is used in order to be able to mark up an inclusion of `postscript` figures, as shown in lines 18-22. The rest should be more or less self-explanatory. The encoded structure for title, author names, bibliography and so on, provide a particular layout format for the document. Equations are encoded as well in the markup language, for example:

Fragment 6:3

```
1  \begin{equation}
2       \bigl[ \pi(\eta,\vec x), \varphi(\eta,\vec y) \bigr]
3       = - i\, \delta^3(\vec x-\vec y)
4       \qquad
5       \pi = a^2\, \varphi' .
6       \label{etcr}
7       \end{equation}
```

...which would render this output:

$$\big[\pi(\eta,\vec{x}),\varphi(\eta,\vec{y})\big] = -i\,\delta^3(\vec{x}-\vec{y}) \qquad \pi = a^2\,\varphi'. \quad (3.4)$$

(*Arbitrary example taken from a preprint on the archive website*)

These examples show the most basic use of TeX markup for the structure and contents of a scientific paper. Scientists can focus on filling in the contents such as `sections' and `equations', although they typically adjusted the markup structure for various purposes. TeX is a programmable environment in the sense that users can configure their own definitions for writing equations and they can include classes and style files to provide additional layout options, hyperlinking functionality, etc. This main source code might also make calls for additional auxiliaries containing, for example, table definitions and table data. Finally, once the code is ready for processing, a TeX processor is invoked and given the main TeX source to produce from this package a device independent output file containing a visual layout description in computer readable code. On the command line, `latex **paperfile**.tex´.

A small mistake, such as "\end{thebibliography}" (fragm.6:2, line 33) accidentally written \end{bibliography}, will halt the processor and feed this message onto the screen.

Fragment 6:4

```
 1 ! LaTeX Error: \begin{thebibliography} on input line 27 ended by
 2 \end{bibliography}.
 3
 4 See the LaTeX manual or LaTeX Companion for explanation.
 5 Type  H <return>  for immediate help.
 6  ...
 7
 8 l.33 \end{bibliography}
 9
10 ?
```

A cursor blinks after the question mark on line 10 and the TeX processor is awaiting further instructions. The processor is configured to communicate what is wrong. It is reading a *closing* mark to match the *opening* mark of the bibliographic section on line 27 in the source (line 1-2).

A logical move is to command the compiler to terminate, open the TeX source in an

editor and fix line 33 to match line 27. But the processor is also configured to suggest reading manuals, which may not be such a bad idea if the screen does not convey an intelligible response, say which is the correct format, `thebibliography' or `bibliography'. Once the lines match, the processor can complete this first stage of building a printable output, `**paperfile**.dvi'. Thereafter, the code is rendered into a postscript by invoking a driver `dvips' and commanding it to produce a postscript output. The new postscript can then be viewed on the screen in a postscript viewer.

This reconstruction shows how a successful correction of a very simple mistake can unfold in the course of producing a scientific document. Scientists can look in the manual or try options that immediately seem logical. In this case, *making efforts* in the use of TeX would be manifested in seeing "Error: \begin{thebibliography} on input line 27 [in the TeX file] ended by \end{bibliography}" on line 33 in the source, "l.33" (line 7), and responding to it by correcting lines 27 and 33 in the TeX source so that the markup matches correctly. But the sequence of successfully accomplishing a markup of a scientific article in TeX can obviously be ordered in a number of different ways and involve contingencies other than this simple reconstruction provides. This work can involve shifting focus, back and forth, between a programmable environment on the display, consulting an available manual and looking at a graphical layout of the output in printable format. There is also an issue of orienting—to know where things are and how they get shuffled around. There could be an issue of using the commands correctly for someone who is new to TeX. The person who is making their "best efforts" would attempt to command a display of built-in help features, look for suitable parameters, test some of them to see what they do, consult a manual, involve someone else, so on and so forth.

A submitter is typically at home in his or her own environment with immediate access to the application software necessary to produce a TeX source and process it locally. If things go wrong the submitter can research the problem and solve it with time and deliberation.

Once a markup and processing of a scientific article is completed successfully on the scientist's local computer, the package can be submitted to the archive. But when the package is sent through the auto-compiler, the submitter is no longer `at home' in the same way and will have to read output logs and warnings, and take corrective action with reference and comparison to goings-on in the local environment. I will come back to examples of contingencies that emerge when using the auto-compiler on the archive but first I will explore the resources for preparing figures.

Towards the bottom of the opening page on the archive website was a link to `available help on submitting and retrieving papers' which took the visitor to an index page with access to information on the submission of figures.

Fragment 6:5

1 Figures for a submission can be in a number of common formats. These include:

2 - PostScript (PS/EPS, most common for line drawings, plots, etc.)
3 - JPEG (ideal for photographic images)
4 - GIF or PNG (best for non-photographic images, bitmaps)

5 [...] If your submission has a large figure, be sure to include it, but make all possible
6 efforts to reduce its size. Note that we do not accept submissions with omitted figures,
7 even if you provide a link to another server. See our information on submission sizes and
8 our bitmapping help for more information.

9 If you submit figures with your (La)TeX source, use standard macro packages (e.g., the
10 `graphics` and `graphicx` packages) so that the figures will appear in the document. [...]

(*From the archive website* [original format])

The figure formats listed with bullet points (lines 2-4) were all the established common standards but the only format that could be inserted into the `TeX` source is `postscript`. The other formats are compromised diminutives of plots or digital photographs, for instance, `JPEG` images are typically lossy derivatives of high resolution photos.

Submitters were advised to produce large figures efficiently and reduce their size if possible, and never to omit any figures from the submission package but rather consult two sources, one about acceptable size of submissions on the archive and the other on bitmapping

figures to reduce their size. Finally, the last two lines advise using the standard `graphics` and `graphicx` packages to insert `postscript` figures into the actual text of the preprint, meaning to mark up their inclusion in the `TeX` source, corresponding to line 2 and lines 18-22 in fragment 6:1.

Two things are of interest here for they often invoked extensive communication with archive operators. One was that plots converted to `postscript` could be too large to be accepted on the archive, even with careful measures to reduce the encoding of them. Conversions into `GIF` or `PGN` bitmaps were necessary. The other issue was the workaround to wrap `GIF` and `JPEG` figures into so-called `postscript` encapsulation in order to embed them in the `TeX` source, but with `PGN` figures this was only possible with intermediate format conversions.

The devices that were suggested to reduce figure sizes, according to help texts on the matter, should be brought into focus. These devices were all freely available on the internet. One example is to use `ImageMagick` on the `Linux` or `Unix` command line, invoking a conversion protocol to go from a `postscript` to a `GIF` or a `JPEG` and back.

Fragment 6:6

- Making a GIF-encoded PostScript:

```
% convert figure.ps figure.gif
% convert figure.gif newfigure.ps
```

- Making a JPEG-encoded PostScript:

```
% convert figure.ps figure.jpg
% convert figure.jpg newfigure.ps
```

(*From the archive website* [original format])

Other suggestions were to use tools to grab an image of a `reasonable graphical quality' off a display on the monitor, and feed it into simple programs like `XV` that offer compressed bitmap formats and reduction of colour maps or conversion to grayscale if there were indeed no colours used in the figure.

To accomplish these tasks was not a straightforward exercise. Reducing file size

inevitably involved some loss of figure quality and the notion of a *reasonable graphical quality* was somewhat problematic. For example, astrophysicists produce enormous photographic images and are reluctant to upload a grainy version onto the archive even if they could leave, in the `Comments' section of the bibliographic data, a link to their own websites with copies in high resolution. But the efforts facing scientists with big figures was that the inevitable loss of figure quality was always a compromise which called for some experimentation: trial and error back and forth, time spent learning about different kinds of imaging software, arguing with the operators and requesting help or dispensation.

To sum up, an oversize problem with figures, faced great variability of formats and file sizes against legibility. There is a myriad of applications available to vary on this theme. But the operators insisted that figure plots in scientific practice can be efficiently and optimally generated with laboratory imaging tools in almost all cases. They also insisted that the engineering environment of imaging should not be much of a mystery to submitters. For example, they could invoke generic algorithms on a `Unix` or `Linux` command line to convert formats, resize, convert colour to grayscale or black-and-white, and more. They ought to be able to see for themselves, with minor efforts, what these tools do by reading about their functions and invoking their help features on the use of parameters. It was also stated in a system-generated email response to processing failures due to oversize packages that the code embodied in a figure file only needed to take up 50 kilobytes of space on average in `postscript` format. Otherwise, as pointed out in the quotes from the help document (fragm. 6:4, 6:5), submitters could try a few workarounds.

Taking corrective action I : lending a pair of eyes

In my own communication with the operators, they insisted that scientists already use `TeX` of some sort to write up their research results and that most of them knew `Unix` and `Linux`. Typically this was indeed the case and evident by simply inspecting the files

submitters were generating. The operators also insisted that scientists are trained to figure things out with a little trial and error. They look for information and want to know the conditions of phenomena. They read instructions meticulously. They read technical manuals all the time. Being trained in scientific computing also meant that they were used to reading all sorts of processing and error logs to figure out problems. The operators recognised nevertheless, that submitters would sometimes ask a question first rather than search through available help texts and guidelines. In that case, the operators would respond with only brief explanation and a reference to the relevant source of information and instruction. In other words, the general message to submitters was to make *more* effort, *look it up yourself.*

Since almost all the submitters were operating TeX compilers and image processing protocols on Unix and Linux, they did not need archive operators to *hold their hands*, as one of the operators put it. A package that compiled and rendered perfectly on a local platform should, in principle, compile just as well on the archive server. If the auto-process on the archive failed in some way however, then the processing log, generated by the TeX auto-compiler, contained error messages indicating what needed doing. The operators also had access to the same processing logs and would often lend an eye to common problems whose source was some error or mistake relating to the relationship between a local machine and the archive server. The operators would point to the source of the problem as a tactic to send the problem back home, so to speak, and remind submitters of available guidelines and instructions including tips about potential discrepancies. Consider this example:

Fragment 6:7

```
 1  Submitter writes:
 2   > ----------------------------------------------------------
 3   > Keyword: "temporary identifier"
 4   > ----------------------------------------------------------
 5   >
 6   > Dear Sir,
 7   >
 8   > I have been trying for more than an hour to submit my paper (manuscript +
 9   > 4 figure files) in the cond-mat archive, all tarred and gzipped into a
10   > single file filename.tar.gz. But everytime I am getting rejected. I have
11   > tried to send the manuscript (in revtex format) and each of the
12   > postscripted figures as separate files too but even that is not proving any
13   > helpful. Since, I have previously submitted numerous other papers in the
14   > archives, I am getting a bit confused at this new turn of events. I will be
```

```
15  > highly gratified if you can kindly help me out.
16
17
18 the error message tells you
19
20 Can't find /home/path/to/figure.eps in ./
21
22 since obviously we don't have your home directory. See the FAQ:
23
24 http://archive.domain/help/faq/mistakes
```

The response begins by stating what “the error message tells” the submitter by quoting the error log: “Can't find /home/**path**/**to**/**figure**.eps in ./” (lines 18- 20). In other words, *this* should have told the submitter that the auto-compiler could not find a figure named `**figure**.eps' in the directory path “/home/**path**/**to**/” because no such path was available in the `working directory', “./”, where the source files were located while they were auto-processed on the archive. The submitter should also know that “obviously” (line 22) the directory path rendered in the warning, “/home/**path**/**to**/”, could not possibly be located on the archive server. This is a directory path on the submitter's local platform and a workable inclusion of a figure can only refer to a figure file in its relevant position to the TeX source holding the inclusion code. If it included reference to a directory structure on a local platform where the package was built, the processing would “obviously” err. The message then points to a website of frequently asked questions where the possibility of accidentally providing so-called absolute file names/paths is mentioned, http://**archive.domain**/help/faq/mistakes (line 24).

What becomes evident in this simple example is the significance of the output log from the auto-compiler which always reported, verbosely, the ingestion and compilation processes step by step. Submitters could consult these log data and inspect any warning or error messages if things went wrong. In other words, the logs were a significant resource to orient submitters to anomalies in the processing and the logs were as well a resource to orient the operators to what had taken place when they were asked for assistance. But building and processing TeX can be expert-intensive and some problems related specifically to the

processing of TeX on the archive server. When this occurred, such contingencies were suddenly very visible—contingencies associated with the maintenance of the auto-compiler. The compiler was supposed to accommodate for many different variations of TeX. There were a number of problems manifested in using it however, all of which pertained to the fact that it was a dynamic and never-ending development project. As I mentioned in the opening chapter, new developments were affected by developments elsewhere such as when new formatting files for TeX markup, classes and styles, were being distributed by journals without securing backward compatibility. The application software had been updated on intervals to reflect new advancements in the markup of TeX. But the most common examples illustrating gradual developments in preparing packages and using the auto-compiler on the archive were manifested in the efforts by the operators to cope with these discrepancies and communicate to submitters established workarounds.

What happened was that some submitters continued to use a range of older and sometimes obsolete components while the new versions were used by default by the auto-processor. Or, the problem was the other way around. Submitters were using new versions of commonly used components while the auto-complier was not yet updated to handle both the old and new versions. This posed a significant problem for the operators because if the auto-compiler was supposed to be updated, it had to be done so that it could proceed without the undesired side-effect that a certain amount of old preprints no longer re-compiled on demand. In these instances, submitters had to be told to include certain class files and other auxiliaries with their packages or add directives. The operators saw these as genuine problems when scenarios in the daily administration reminded them that further development work was needed. Submitters were typically not aware of these discrepancies until they ran into a halt and contacted www-admin because of some warning or a rejected submission.

Fragment 6:8

```
1  […]
2   > Please help me, what is the problem? Is it related with Latex2e?
3
```

```
aipproc.cls has changed too many times without any name change. You have
to bundle your version of aipproc and the necessary auxilliary files
with your submission.[…]
```

The American Institute of Physics (AIP) journals had been distributing a new layout package with a backward incompatible class, `“aipproc.cls”`, and the submitter is unsure of what exactly was causing the process to halt, `“What is the problem?”`. The operator could assume that this same source compiled perfectly well on the submitter's own machine where another version of a class file named `“aipproc.cls”` and other auxiliaries from the publisher were installed. The matter is explained, although very briefly, assuming that the submitter can see precisely what `“bundle your version of aipproc and the necessary auxilliary files”` means in terms of how to proceed.

A few problems of this sort recurred often enough and had been around for long enough to justify a special `frequently asked questions' (FAQ) page.

Fragment 6:9

```
see

http://archive.domain/help/faq/amslatex2000

[…]

 > The file is in standard "amsart" and runs fine with our local latex
 > installation (no error messages whatsoever), but leads to an emergency
 > stop when submitted to the […] preprint server, where the core of the
 > error message seems to be the following:
 >
 >   […]
 >   LaTeX Warning: Citation `FI' on page 1 undefined on input line 129.
 >
 >   Runaway argument?
 >    \noindent We shall call locally conformally parallel $\mbox {G}_2$-
 >    m\ETC.
 >   ! Paragraph ended before \equation was complete.
 >   <to be read again>
 >                       \par
 >   l.309
 >
 >   ?
 >   ! Emergency stop.
 >   <to be read again>
 >                       \par
 >   l.309
 >
 >   No pages of output.
 >   Transcript written on mathpaper.log.
 >
 > We reinspected the lines around l. 308 carefully for syntax errors,
 > but could not find any mistakes in the vicinity of these lines.
[…]
```

In this case, it is a class from the American Mathematical Society `amsart' which has caused the problem. The source called for the wrong version of it and did not compile. The auto-compiler has communicated where the process comes to an `"emergency stop"` (lines 15-30) but I will first draw attention to lines 7-8 and 32-33. The submitter claims that the `"file runs fine with our local latex installation (no error messages whatsoever)"`, then claims to have `"reinspected the lines around l.[line]308 carefully for syntax errors"` without finding any mistakes in the code (lines 32-33).

From these claims it is clear that the processing log on the submitter's end was not reporting the warning on line 13, nor the emergency stop quoted on lines 15-30. There were `"no error messages whatsoever"`. On the archive, however, a warning message from the auto-compiler first reports an undefined citation on line 129 in the code and then halts when the compiler detects an end of paragraph before an end to a markup of a formula on line 309. It asks if there is a `"[r]unaway argument?"`. But rather than explaining in detail why this happened, the operator points to a help text or a FAQ page, explaining how to use `amsart` on the archive. The first remark in this message `"[t]he file is in standard "amsart" and runs fine"` confirmed to the operator what the problem was. The application software on the archive had not been updated to compile `TeX` using what was the latest standard `amsart` document class.

From these examples, we learn that even to the `TeX`, `Unix` and `Linux` literates, the archive was not entirely a self-service. One could of course argue that in these particular examples submitters could have sorted matters out by themselves with *more effort*, a little bit of searching in the help texts on the archive web site, `http://archive.domain/help`. But they did ask questions and they did receive answers, albeit succinct, eg, `"aipproc.cls has changed too many times without any name change […] bundle your version of aipproc and the necessary auxilliary files"` (fragm. 6:7, lines 4-5). and an access link to an

information page on what was the case and what to do about it, "See the FAQ: http://**archive.domain**/help/faq/mistakes" (fragm. 6:6, lines 22 and 24) or "see http://**archive.domain**/help/faq/amslatex2000" (fragm. 6:8, lines 1 and 3). I will return to these conventions in the discussion section but now turn to more elaborate efforts by submitters and operators to communicate procedures. I will enquire into communication relating to the preparation of figure files.

Taking corrective action II : communicating procedures

The diversity of image formats and image processing for all sorts of purposes and for different platform environments is such that it renders it impossible to know everything there is to know about it. However, the most common devices to post-process figure plots among the registered submitters were run on Unix or Linux platforms. As we learned in the section on resources, submitters were told to generate figure plots efficiently, to compress the figures adequately or bitmap them in order to keep the total size of the submission package below a certain size limit. Packages that exceeded the limit were automatically rejected and submitters would either have to make efforts to compress the figures further or they contacted the operators and requested dispensation.

We learned in the previous chapter (fragm.5:12) that if the number of figures was high, and figure generation was determined optimal on inspection, the package would be manually inserted into the pending queue on the day. The submissions that required more work, however, are of some interest here because the operators often provided considerable support. We shall then look at an example of how this could play out.

In the following sequence, the owner of an auto-rejected package communicates:

Fragment 6:10

```
1 […] i would like to put on the archive lectures notes. Because these
2 lectures notes contain 43 figures their size exceed the standard size of
3 the archive. I have done my best to compress the figures (most of them
4 are now below the 50k limit, and the biggest one (experimental image) is
5 268k). Because of the numbers of figures it is unlikely that much further
6 compression could be acheived. […]
```

The operator who caught this message grabbed one of the figures from the temporary holding and reprocessed it. The reply then offers the submitter to check the difference by accessing both the original and the new `postscript` through a web page.

Fragment 6:11

```
 1 […]
 2 Compare the following two versions of your figure:
 3
 4  http://archive.domain/figure.eps.gz
 5  http://archive.domain/new-figure.eps.gz
 6
 7 The first is the 214kb file that you sent. The second was created here
 8 and is a 74kb file, only a third the size.  If you agree they are
 9 equivalent in quality then your gzipped [compressed] submission can be
10 reduced dramatically in size without loss of image quality.
11 […]
```

This reply does not tell the submitter how the second figure was created, nor did the submitter provide information in fragment 6:10 on how the original was produced. But the two links (lines 4-5) provided access to comparison between the original and the image compressed to 74 kilobytes (lines 7-8).

What was actually done on both sides was uncertain at this stage but became clearer in the following two messages. First an excerpt from the submitter:

Fragment 6:12

```
 1 […] Some of the figures i sent were created by passing the much larger
 2 scanned images given by my co-author through the acrobat distiller, then
 3 converting them to postscript again (this was the case of the example
 4 image you used).
 5
 6 The quality of the smaller image you sent is certainly satisfactory, so
 7 if you can automatically compress the other images as well and proceed
 8 with the submission, this will certainly be fine.
 9 If not please let me know the procedure you have used (this could be
10 useful to know in any case), and i'll try to further compress the images
11 […]
```

In the first paragraph (lines 1-4), the submitter only partially explains how the figures were generated. Scanned images from a colleague were exported through the `acrobat distiller`, a driver that creates `PDFs`. Then the `PDFs` were converted into `postscript` encapsulation. Neither the format of the original scans nor the method of conversion from

PDF to a postscript is communicated but, in the second paragraph, the submitter suggests that "if you [admins] can automatically compress […] and proceed with the submission, […] fine […] [but otherwise] please let me know the procedure you have used". There could be something useful to learn from it (lines 7-10).

As it turned out, there were no automated processes to invoke on the back end in order to convert and compress batches of figures. As we have already learned, submitters were told to make all possible efforts, "If your submission has a large figure, be sure to include it, but make all possible efforts to reduce its size" (fragm. 6:4, lines 5-6). The submitter also adds, "let me know the procedure you have used […] and i'll try to further compress the images". The reply from the operator responds specifically to this particular request.

Fragment 6:13

```
 1 […]
 2 > converting them to postscript again (this was the case of the example
 3 > image you used).
 4
 5 Highly inefficient. To create sample, the [your] postscript was
 6 converted to greyscale png using xv, then using the ppm tools
 7 pngtopnm figure.png | pnmtops -rle -noturn > new-figure.eps
 8
 9 But it is better to start with the scanned images directly. Depending on
10 format you could substitute tifftopnm or giftopnm in the first step
11 above, e.g.
12 giftopnm figure.gif | ppmtopgm | pnmtops -rle -noturn > new-
13 figure.eps (where ppmtopgm converts from color to greyscale to remove
14 color scanning artifacts in the case of b&w images, and the -rle for
15 "run Length encoding" is suited to scanned line drawings).
16 […]
```

The operator first replies directly to a paragraph copied in from the earlier message (lines 2-3) by commenting succinctly on the description of efforts as "[h]ighly inefficient" (line 5), and then proceeds to communicate how one of these *inefficiently* generated postscripts were further processed on the archive (lines 6-7). To explain, the operator used the figure processing tool, XV, for Unix and Linux, as well as the generic tools, PPM. First the grayscale figure, in full colour-encoded format, was "converted to greyscale png using xv" (line 5-6). Thereafter the grayscale PNG figure was sent through two conversion tools in one

command to produce a new postscript output, using the PPM tools. The full command is given on line 7 as,

```
pngtopnm figure.png | pnmtops -rle -noturn > new-figure.eps
```

Briefly, the tool, pngtopnm (literally PNG to PNM), is invoked and given the PNG figure, "**figure**.png", for processing. It converts the figure into a so-called `portable anymap' image (PNM) and then this PNM output is directly `piped' into another conversion tool, "| pnmtops", along with a couple of parameters, "-rle -noturn", and a postscript output command with a placeholder, "> **new-figure**.eps".

What is visible in these first three lines (line 5-7) is how the operator communicated a particular process on the assumption that the submitter was familiar with the Unix or Linux environment, and with command line arguments and parameters that activate and control figure processing tool. The new figure was created "using xv" and "using the ppm tools". This process involved using the command "|" to `pipe' the figure through the pnmtops algorithm. It involved using the command ">" to open a placeholder for a specifically named postscript output. It also involved using two parameters, "-rle" and "-noturn", each of which provide directives to pnmtops on how it should proceed. The reason for using these parameters is briefly explained on lines 14-15. The submitter should see that the former, "-rle" or "run length encoding" (line 15) runs a compression that collapses many consecutive data elements holding the same data value and mostly found in line drawings. The submitter should also see that using the latter, "-noturn", will command pnmtops not to automatically turn a figure which is wider than its height.

Lines 9-14 of the last paragraph continue on a similar note to lines 5-7, providing further advice on how to achieve efficient results directly from a plot output either in TIFF or GIF format. This is manifested first in the remark that "[d]epending on format you could substitute tifftopnm or giftopnm in the first step". The sample command is formulated accordingly as an example of compressing a plot in GIF format. The figure is

converted directly to `PNM` and then an additional `piping' is inserted into the command, `"| ppmtopgm"`, with the explanation that `"ppmtopgm converts from color to greyscale to remove color scanning artifacts in the case of b&w [black and white] images"` (line 13-14). The communication ended here and the submission was uploaded later that day. An advice had been provided on how to do this efficiently and the next move was thus with the submitter to get *better* results with *more effort*, perhaps with some trial and error in coping with the contingencies particular to figure processing with `Unix` and `Linux` tools.

Taking corrective action III : removing preprints

One thing we have learned so far is that disorderly preprints might simply be removed from the system with a brief message to the submitter on what to do next. When this particular action was taken, the responsibility to *take corrective action* was sent home with considerable force, ie., no posting on the archive. Two scripts were used by the operators to do this, `` `lastmv' `` and `` `lastrm'. `[L]astmv' `` was written to remove a package from inside the pending queue but in order to keep the serial numbering intact, the script also routed the very last incoming package into the place that opened up. `` `[L]astrm' `` was written to be invoked only when the very latest package was removed. One was therefore concerned with a single package while the other dealt with two packages.

We learned in the previous chapter that `awareness data' on package relations and dependencies were generated by the apparatus and distributed to databases inside the server machine. Thus, removing a package or moving it between subject archives required using these scripts. They were written to also take care of these data, removing them or altering them as needed.

To know exactly what `` `lastmv' `` and `` `lastrm' `` actually did, one would have to read them carefully. There was an incident on the archive however, which brought sharply into focus what these script do, ie., what would happen if one of them was *not* used to remove a

preprint. It was a mistake in command by a new assistant techy. Three operators were involved in this scenario, **AS** (the assistant techy), **AA** and **AB** as well as the system **SYS** and a submitter **S02**.

Fragment 6:14

15:32 **AS**: to **S02** (copy ***to admins***, subject: [archive] submission rejected (**archive/papernum**))

> We do not accept low quality PS using bitmap fonts.
>
> Your submission looks as if it was produced from tex. If that is the case then you must submit the tex source. See: http://**archive**.**domain**/help/faq/whytex.html

15:55 **SYS**: ***to admins***. (subject: final freeze status (5 minutes to go!))

> […]
> **archive/papernum**: Error - can't open /users/**admins**/**archive**/papers/0111/**papernum**.abs : No such file or directory at /home/e-prints/bin/epcat line 10.
> […]
>
> Parse errors
> **archive/papernum**: >>>> FAILED TO PARSE ABSTRACT <<<<

16:14 **AA**: ***to admins*** (*subject*: freeze problem)

> [quotes system diagnosis …]
> /users/**admins**/**archive**/papers/0111/**papernum**.abs: No such file or directory
> SERIOUS PROBLEM (FIX MANUALLY): **archive** zero size **papernum**.abs
> [end quote]
>
> who moved **archive/papernum** ?

16:16 **AS**: ***to admins***. (*subject*: **archive/papernum**)

> [quotes **SYS** 15:55]"
> **archive/papernum**: Error - can't open /users/**admins**/**archive**/papers/0111/**papernum**.abs : No such file or directory at /home/**admins**/bin/epcat line 10.
> "[end quote]
> ???
>
> This is the paper I removed around 3:30. [see lines 1-8]

16:17 **AA**: ***to admins***. (*subject*: **archive/papernum**)

> **AS**,
>
> did you see any error message when you removed this?
>
> Did you do a lastrm or a lastmv?

16:30 **AA**: ***to admins*** (*subject*: Re: **archive/papernum**).

> **AS** writes [to **AA**]:

```
58             > my fault, since it was the last file in i did an 'rm' from the
59             > directory.
60             > i didn't know about lastrm. Sorry.
61             > […]
62
63            ok, you can't simply remove papers.
64
65            there are password entries, title databases, counters, etc. to
66            reset, too.
67
68            always use a script
69
70            […]
71
72            This is simple to fix, no big deal, however requires some careful
73            checking, will take care of it.
74
75
76  16:56     AB:  to admins (subject: Re: freeze problem)
77
78            this is completely incredible to me...
79            what about the running papernum, title database, password
80            database, etc???  […]
```

The two scripts `lastrm' and `lastmv' are referred to by **AA** on line 52. The fact that both scripts took the necessary care of password entries, title databases, counters, and numerous other records of processed packages, is specifically stated here by **AA** and **AB** (lines 65-66 and 79-80) in response to the use of `rm' which is the device **AS** had used (line 58). `[R]m' is a generic Unix/Linux device invoked on the command line to delete or `remove' one or more files, eg., `rm file1 file2 file3'. In this case, a data package and the bibliographic record for it were removed with `rm', but the system still attempted to parse the record, `**papernum**.abs', gather all information on the package, detect and harvest anomalies. So, when system awareness of this and all other new packages on the day was invoked at 15:55, the processing errs (lines 14-20). I will now take a closer look at this sequence.

This package was not reported in the standard fashion in the `prefreeze check' report at 15:15. The only explanation for this is that the package had not yet been submitted. The only record of this single postscript, before 15:32 when the sequence begins, is a metadata report among other reports on new submissions that **AS** was attending to on this day. The postscript would have been inspected by **AS** because **AS** removed it on the grounds of being "low quality PS", looking "as if it was produced from tex", as stated in a canned

message to the submitter at 15:32. (lines 4-6). Then at 15:55 the system sent out its `final freeze status' report with two error messages (lines 14-16 and 19-20).

The problem was first communicated among the operators at 16:14 when **AA** quotes further system diagnosis of a freeze problem (26-28). "**[P]apernum**.abs" could not be found and the system was configured to recognise this as a "SERIOUS PROBLEM" and command the operators to "(FIX MANUALLY)" (line 28). **AA**'s first question was, "who moved **archive/papernum** ?" (line 31). Another two messages were sent into the message box in the next three minutes. First **AS** quotes one of the error messages from the `final freeze status' report (lines 37-39) and adds three question marks "???" (line 41), indicating confusion but then adds, "[t]his is the paper I removed around 3:30" (line 43). Next, **AA** addresses **AS** directly with two targeted enquiries. Did **AS** "see any error message" and was **AS** using "lastrm or a lastmv" (lines 50-52)?

A further exchange between them becomes visible at 16:30. **AS** had contacted **AA** directly with answers, bypassing the message box. **AS** "did an 'rm' from the [packages] directory" and did not "know about [the script] lastrm" (lines 58-60). This particular message is then included in **AA**'s open response to **AS**. First there is a gesture of realisation, "ok" (line 63), and then **AA** explains why a package cannot be deleted like that, "there are password entries, title databases, counters, etc. to reset, too" (lines 65-66). A operator should "always use a script" (line 68). Then the involvement of **AS** is closed with a statement that **AA** can fix the problem and will take care of it, although it "requires some careful checking" (lines 72-73). **AC**'s input in lines 78-80 frames how serious this breach is, and is interesting in its own right, however, I wish to focus on the work of identifying critical resources to establish intelligibility and accountability, ie., finding or seeing what has happened and what to do next.

When **AA** noticed that the system reported, "SERIOUS PROBLEM (FIX MANUALLY)" and that a bibliographic component of a package was missing, efforts to cope and cooperate had a

paramount role in sorting the matter out. As in the example of commanding `pscheck', which found no match, communication between operators and computer supports breaks down and needs to be restored. In this case, the system was configured to signal a performance-critical state, a "SERIOUS PROBLEM". Expected data were missing in a data stream which the system was designed and configured to handle in very specific ways. It was not immediately evident, however, how to competently engineer a recovery. The computer supports at work in the background took the first step by detecting a problem, flagging a warning and commanding investigation, diagnosis and a fix. **AA** sees what to do next, ie., to probe for cues to establish an intelligible account of what happened and what the current status of the system is. But this work calls for cooperation which can be summarised as follows:

AA Who moved the package?
AS *I removed it.*
AA Did you see any error messages? did you use, `lastmv` or `lastrm`?
AS *I used* `rm`.
AA `ok`, this is what happened and here is why to use our own scripts, and now I have targeted what needs to be done.

First, **AA** wants to know whom to address: *who is in here?* Secondly, **AA** probes directly for evidence that might contain target-specific information such as which device was used and if it reported anything, *what did you see, what did you use?* The reference to `rm' is the key here to clarifying what requires fixing. It is an instant explanation and, consequently, what needed doing was immediate manual realignment of certain computer-generated data relations and dependencies.

Discussion

Now that I have laid out examples of taking corrective action, I will look closer at the ways in which effort-making and competence are relevant to doing this work. In the first section, we examined how submitters are addressed to resolve problems before seeking help from an operator. The auto-response (fragm. 6:1) communicates the efforts they should make

and whom to consult before sending an email to `www-admin`. It gives a sense of what is expected of them and the initiatives they should employ to solve problems, ie., `"a little trial and error"` or consult `"a computer savvy colleague"`.

By illustrating a simple reconstruction of how TeX is prepared and processed (fragm. 6:2, 6:3 and 6:4) we learn that TeX is a tool box and a host of building blocks, and it requires a bit of tinkering to mark up an article. Users of TeX employ a range of styles and classes and they improvise with definitions, formulae, table data, hyperlinks and more. In other words, marking up in TeX is not exactly a straightforward and strictly ordered exercise, neither is the preparation of figures. When we consulted available help on preparing figures (fragm. 6:5) we learned that all possible efforts need to be made to reduce them in size. Figures are plots, line drawings or photographs, in a range of formats, some of which need to be manipulated using conversion and compression tools and then encapsulated in postscript for inclusion in TeX. What we see then is that expectations are communicated to submitters that they make 'their best efforts', even though the practical use of TeX and figures in scientific practice requires considerable competence in using the available tools and supports.

I have already argued that what we can see of the operation and use of the archive does not have a place in cognitivist reasoning about isolable agents and input/output processing or formal explanations of cognitive processes interacting with information processing software. I discussed some of the alternatives to that line of reasoning in Chapter 3. For example, so called `post-cognitivist' theories, activity theory and the theory of distributed cognition (DC), emphasise the social, cultural and historical contexts of doing and organising work. Both theories acknowledge that actions and interactions are situated encounters. However, DC theorists rely on normative accounts to stabilise the structural properties of work and its organization, while activity theorists insist that motives and goals drive people in their work and thus determine how it can be supported and organised. We also learned in Chapter 3 that cognitivist reasoning is critiqued by sociologists, anthropologists and philosophers for its

depiction of disembodied knowledge. For example, sociologists draw attention to socially constituted attributes grounded in social epistemology, however, such attributes tend to be cast in very generalised terms. Knowledge and skills are *practices* which are socialised, enculturated, discursive or embodied. Correspondingly, practices of interacting with complex objects, instruments and environments are what persons *stand in* for when they are described as technical experts or competent technical workers. The objection to these lines of reasoning is that in reducing knowledge to social interactions, socio-material practices or embodiment, the attention is shifted away from the actual details of lived working practices—how people rely on and use tools and supports to get on with work, cooperate and organise procedures. Such details are also neglected in designs of computer supports that seek to capture the relevance of specific tools and supports in the context of motives or circumscribed goal-oriented tasks.

Protocol standards, compatibility, interoperability or specialised purposes are all practical matters taken into account when choices are made about document and information handling. I pointed out in Chapter 3 that studies of those who use open source software and coordinate operations in virtual spaces teach us that basic protocols and generic tools are preferred. They can be improvised in how they are used and assembled. Strictly purpose-built designs do not give the necessary flexibility to accommodate options that are immediately relevant in some particular case. As I explained in Chapter 1, the archive apparatus is an amalgam of computerised document/communication/operations/technologies. What we have now learned is that using TeX and manipulating figures is faced with assemblies of generic tools and basic protocols to adequately build a preprint for a submission on the archive. So what counts as competent use of tools and supports to keep preprints in good order?

For example, in fragment 6:9, we see how the submitter makes visible relevant items of information when attempting to submit a preprint to the archive. The submitter offers an excerpt from an error log containing `"! Emergency stop"` (line 24) of the compiler and on what line in the code the compiler halts. The submitter is also particular about encoding the

details of what he has done and what he has seen, “we reinspected […] carefully […] but could not find”, pointing specifically to the vicinity of where in the code the processor halts. In other words, the submitter cannot find an error but orients the operator to the exact location reported by the compiler. In the first paragraph of the message, the submitter makes the point of naming the class of the file, `amsart', and stating what can be eliminated, “runs fine with our local latex installation (no error messages whatsoever)”. In this statement, the submitter is carefully drawing attention away from questions of their own competence, while persuasively suggesting the place where inspection is needed. Not knowing a particular contingency in using `amsart' *on the archive* is a separate matter identified by the operator.

In fragments 6:10 to 6:13, we see evidence of elaborate interaction between submitter and operator. In the first of these the submitter makes the point of stating “I have done my best”, explaining that his submission has 43 figures “most of them are now below the 50k limit”. In this message the submitter is asking for dispensation on the grounds that the number of figures is high and “unlikely that much further compression could be achieved”. The operator who responds to this message in fragment 6:11 makes readily available two versions of the *same* figure, one from the original submission and the other recompressed by the operator, “only a third of the size” to demonstrate that further compression *is* possible and the submitter can now assess whether or not the figures “are equivalent in quality”. In this exchange, we see how the operator orients the submitter to evidence that, despite the submitter's *best* efforts, not enough effort has been made. The submitter's response to this, in fragment 6:12, is to list the tools and procedures used to create “some of the figures”, then offer an exchange of information on tools and procedures, ie., “the figures I sent were created by... […] please let me know the procedures you have used” (lines 1 and 9). When the operator replies the second time, a detailed description of procedures displays the names of the tools that were used and two full commands to

exemplify how figures *can* be compressed more efficiently.

Built-in configurability and flexibility of tools and supports is also manifested in reference to a number of different devices and specific mention of their uses (fragm 6.12, 6.13 and 6:14). They are open to experimentation, ie., trial and error. Issues of competence and *making efforts* are made relevant insofar as submitters and operators cooperate using mutually familiar terms. We learn here is that competence is not *defined* by protocols and methods that need to be known in order to do certain tasks. Rather, it is established and negotiated in the exchange of efforts which result in seeing how to move forward in the most effective or efficient manner.

In this chapter as well as Chapter 5, we see how *looking for and seeing* takes precedence in getting on with the work. Inspection and detection is an integral feature. This particular observation corresponds with studies in ethnomethodology that demonstrate the significance of visibility, location and ways of orienting the attention of oneself and others. Submitters and operators name the tools they use and the phenomena they see and want others to see. They orient themselves and others to exact locations by pointing and naming: it is *here* and *there* that one should look at *this* and *that*. To follow Lynch's argument (2006), there is site-specificness to the singular performance of looking for and seeing, and there are contingencies in *doing* just that. We see deliberate and selective efforts to make precise task-relevant information visible for particular purposes at hand. This is already given considerable attention, for example, in Health and Luff (2000).

This matter is starkly emphasised in the last scenario (fragm. 6:14). The example I took concerned the use of specialised devices to remove preprints and an incident of using neither of them. The operator who is alerted by the system to a "SERIOUS PROBLEM" *demands* the naming of the tool used by **AS** and *demands* to see what **AS** saw. In other words, the sequence of clarifying how a recovery can be engineered is initiated through a series of probes: *who* did this, *what* device was invoked, *what* did you see. The openness and thus perhaps vulnerability of an open and programmable environment was suddenly visible when a

mistake in command brought the whole apparatus to a performance-critical state. But what 6:14 also teaches us is that while the operators are isolated, each with a monitor and keyboard access to the apparatus, they keep a check on each other and the tools and building blocks they work with in ways that selectively and meticulously provide visibility. An attempt by **AS** to communicate privately about what happened is responded to by swiftly inserting that communication back into the message box for *all* operators to see. And, despite the necessary relay in asynchronous communication, or rather because of it, **AA** commands coherence and continuity by first *demanding* adequate visibility for orientation purposes.

In the next chapter I will discuss procedures of preparing and submitting preprints files which were much less obvious and accessible, both to operators and submitters. I will also address problems of integrating library personnel into the daily administration of preprints. These events raised issues about the usability and usefulness of tools and procedures.

Chapter 7

Usability and usefulness

Introduction

The previous two chapters enquired into the ways in which inspection, detection, and taking corrective action was carried out—work which, in one way or another, involved problem detection and problem-solving efforts. This chapter will continue this line of enquiry with examples where the expectation of those who were involved in doing this work were challenged. In other words, this chapter will take examples of how operating and using the archive came under scrutiny. The scenarios I will consult and reconstruct relate to changes in the way users were preparing their submission files. They were using graphically interfaced applications. I will also consult scenarios relating to changes in the daily administration of preprints when members of library staff came on board. It was unclear what skills were required and what alterations were needed to the administrative set-up.

The archive apparatus, as I came to know it, was open-ended with an uncertain future. The operators treated the archive as a laboratory for information and document handling technologies, ie. apart from administering new preprints on a daily basis. The invention of this `self-service' did not originate in software design and engineering aimed at a neatly packaged and user-friendly protocol for general self-archival purposes—an `off-the-shelf' product that did not require expert-intensive support and software to successfully process TeX. Obtaining the source code of preprints, unless such a source was not produced at all, was seen to be the guarantee of article availability in the future regardless of how information and document handling technologies might develop.[9]

[9] For example, it was not clear at all in 2001 and 2002 if the PDF file format would prevail and there was no way of reverse engineering PDFs, although, later this has become possible to some extent.

The operators were also manifestly opposed to the use of applications that only run on Microsoft or Macintosh platforms. Microsoft products, in particular, were snubbed. A couple of operators habitually wrote wind(bl)ows, micro$oft or m$word, if these products had to be named—adding cynical references to user ignorance and cost as opposed to for-free, open source and expert-intensive products like those used by the operators and a vast majority of submitters. As we have learned, registered submitters were generally trained in the use of Unix and Linux, but many were also active contributors to the development of software that runs on these platforms. *For-free* and *open-source* are often mentioned as the qualities that attract scientists to use these platforms. There is a significant practical advantage – to be able to know what a piece of software is actually doing by inspecting code as well as logs. Unix/Linux environments are wide-open to investigation and engineering. However, if the output from a proprietary software is subject to scrutiny, or if there are problems with the functions of proprietary platforms, the user faces a `locked box'. Output logs are not very instructive of what the software is actually doing nor does the user have access to any source code.

In this chapter, I will enquire specifically into issues of usability and usefulness. Trends were changing in the preparation of preprints and conditions were changing in the daily administration of them. I will enquire into communication with submitters about preprints prepared with software built for Microsoft and Macintosh platforms and I will enquire into manifestations of unrest about what the administrative work required.

Issues of usability and the usefulness of available computer and expert supports, raise further issues of relevance. On the one hand, questions were raised about learning and adaptation manifested in the work of bridging, or attempting to bridge, significant procedural gaps. On the other hand, communication on how to move forward with the administrative work raised questions about access to the actual circumstances of doing the technical work.

In the next five sections, I will consult and reconstruct scenarios to shed light on these

matters. The first two concern the use of software running on Macintosh and Microsoft platforms to prepare either TeX or figures. The next two sections concern episodes of planning `talk', one between the operators and library staff, and the other also involving senior managers. Plans were made on moving forward to reorganise the administrative environment and create labour divisions. Plans were also made for library staff to join the daily administration of preprints. The section on using the daily grind will then consult a recording of myself explaining some of the administration work while doing it.

In the final section, I will relate these scenarios to issues of usability and usefulness. I will discuss the unrest with procedures as well as the ways in which issues of useful and usable computer supports take on their practical relevance in relation to design considerations.

Using TeX

Submissions produced with Scientific Workplace for Microsoft Windows were relatively rare. They were not supported specifically by the TeX auto-compiler during my time on the archive but the software could export to a standard LaTeX source. These exports would normally process on the archive as long as software-specific class and style files, or other auxiliaries were included with the package, ie., if their inclusions were also embedded in the code of the LaTeX source.

Communication with submitters who used this graphically interfaced application was mainly about the inclusion of figure files. One problem was that figure files would have names in upper case while the inclusion code referred to the file names in lower case. Submitters were often not aware that Unix/Linux is strictly case sensitive. The exchange I will consult, however, concerns a problem of how to use certain text encoding. This exchange takes place on two days involving an operator **AA**, a submitter **S03**, and, in italics, a quote from a message from the technical support at Scientific Workplace.

Fragment 7:1

7 Nov
18:44 **S03**: ***to admins*** (*subject*: RE: preprint upload impossible in TEX)

[…] I contacted the Scientific Workplace people, as you suggested, and, after much back and forth, they seem to conclude that

> *The LaTeX format file used by the automatic system uses OT1*
> *encoding, but T1 encoding is needed to use \dh [ð]. I searched*
> *the archive server but do not see that it is possible to use a*
> *different format file. I think you will have to rewrite your*
> *document without using the edth character or find out how to use*
> *a format file with T1 encoding.*

I cannot do without the edth [ð] character in my paper. It is absolutely essential. I therefore request that you let me upload a postscript version of the paper.
[…]

18:55 **AA**: ***to*** **S03** (copy ***to admins***, subject: RE: preprint upload impossible in TEX)

\usepackage{t1enc}

the sciword tech support should know at least the standard latex options, this is part of the base distribution. There are many other ways to select t1 encoding, […]

9 Nov
14:04 **S03**: ***to admins*** (*subject*: RE: preprint upload impossible in TEX)

So what do I do with this command: \usepackage{t1enc} ?

Why won't you let me use postscript? This is getting ridiculous.[…]

14:09 **AA**: ***to*** **S03** (copy ***to admins***, subject: RE: preprint upload impossible in TEX)

insert this command into the preamble of your latex source.

14:29 **S03**: ***to admins*** (*subject*: RE: preprint upload impossible in TEX)

I did. It still doesn't work (**temporary identifier**). What now?

14:37 **AA**: ***to*** **S03** (copy ***to admins***, subject: RE: preprint upload impossible in TEX)

this should be obvious:

```
   ! LaTeX Error: \usepackage before \documentclass.

   See the LaTeX manual or LaTeX Companion for explanation.
   Type  H <return>  for immediate help.
    ...

   l.2 \usepackage{
                  t1enc}
```

The problem communicated by the submitter concerns the rendering of the character "\dh" or `ð' as it would look like in the preprint. The submitter claims to have "contacted the Scientific Workplace people" (line 4) and quotes their response to the problem (lines 7-12). They refer to a "format file" on the archive system, using "OT1 encoding" whereas "T1 encoding is needed" (lines 7-8). They do not "see that it is possible to use a different format file" as they put it (line 9-10). They suggest to "rewrite [… the] document" omitting the character (lines 10-11). The submitter, however, "cannot do without the edth [ð] character in [… the] paper" (line 14) and asks permission to "upload a postscript version" (lines 15-16). This is not an option on the archive.

To explain, the answer to this problem would not have been found on the archive web pages. It is in the LaTeX manual. The issue is not "to use a different format file" *on the archive*. Rather, the TeX source should be differently formatted so it invokes the use of another text encoding package. The missing command is given away on line 23, "\usepackage{t1enc}", but lines 26-27 also state that "[t]here are many other ways to" do this.

Although the submitter is using a graphically interfaced software, the TeX source output is accessible and can be edited directly as plain unformatted text. Two days later the submitter asks, "[s]o what do I do with this command: \usepackage{t1enc} ?" (line 33) and the answer, a few minutes later, is to "insert [it] into the preamble of [… the] latex source" (line 41). The term, `preamble', is named specifically by the operator as task-specific. The command should be inserted with other directives in the first lines of the TeX source where declarations of document class is put, what packages are needed or mathematical definitions, ie., *before* the standard \begin{document} mark (see Chapter 6, fragm. 6:2, lines 1-2). Shortly, the submitter contacts the administration again, claiming that "it still doesn't work […] what now?" (line 46) and receives a prompt answer, remarking that "this should be obvious" (lines 52). The submitter would have to see what the TeX

compiler is communicating, "[The e]rror: [is that the directive] \usepackage [comes] before \documentclass" (line 55). The compiler also suggests, as it routinely does, to consult manuals (line 57). In short, a TeX complier first needs a declaration of the document class it has been invoked to process *before* it can be commanded to use certain packages, definitions, and so on.

Regarding this preprint, the submitter had claimed earlier not to know any TeX and in this sequence repeats the request to upload a postscript output from Scientific Workplace (lines 15-16 and 35). To align the TeX output with the auto-compiler takes resources but the resources made available by the operator and the compiler on the archive are not immediately usable or useful. They are neither familiar nor ready-to-hand, "what do I do […]?", "[i]t still doesn't work", "[w]hat now?" (lines 33 and 46). The submitter is moving within an environment locally which does not provide an intelligible comparison or bridge with what happens on the archive. The command, the information on where to insert it, and the error log, do not convey intelligible information. There is no indication either that a TeX manual has been consulted.

The incident brings home with considerable force the pressure to learn about preferred objects / devices for using the archive: how to control them and read their outputs. Both requests to submit a postscript are ignored. But one can argue that an orientation to technically uninformed and graphically interfaced use of software is not easily consolidated with an orientation to open machine and software code. The operator has nothing to offer in this exchange that is specific to the use of Scientific Workplace and makes a remark to the effect that openly disapproves of the technical support personnel for not understanding basic LaTeX markup, "sciword tech support should know at least the standard latex options" (line 25). Under the circumstances, however, that T1 encoding is needed to render this preprint adequately, the operator names phenomena relevant to the handling of LaTeX.

S03 I cannot do without the edth [ð] character
AA **\usepackage{t1enc}**
S03 what do I do with [it …]?

AA insert […]into the **preamble** of your **latex source**
S03 [i]t still doesn't work
AA **Error: \usepackage** before **\documentclass**

Bridging gaps in understanding what to do and making use of the auto-compiler is manifested in these particular specifics of the exchange. But these are, at best, fragmented attempts to align *sufficiently enough* in order to move on, rather than a concerted attempt at mutual alignment and understanding. I will now turn to scenarios involving submitters and their figures.

Using figures

During my time on the archive, submitters had sometimes imported large digital photographs or vector graphics into imaging software from Adobe Systems, `Illustrator` or `Photoshop`, to produce `postscripts`. These software are graphically interfaced and run on `Macintosh` and `Microsoft Windows` platforms. It is quite easy to learn how to invoke `` `import' `` functions and load graphics onto the screen where they can be looked at and checked, perhaps resized by using the `` `image' `` options and then saved or printed as `postscripts`. These processes are invoked by the click of a mouse and the interface is not expert-intensive in the sense that options and settings are laid out graphically and without direct access to the underlying processing environment. With a moment of deliberation, perhaps consulting a quick-tips manual, trying a few options, and so on, producing a `postscript` can be learned quickly and intuitively.

Since the operators did not operate `Macintosh` or `Microsoft Windows` platforms, they did not run `Illustrator` or `Photoshop` either. A letter might come into the message box asking for dispensation or advice because of an auto-rejected oversize package. The operators would inspect the figures on the back end, for example, look at their contents in text view. In the case of an `Illustrator` or `Photoshop` generated `EPS`, the operator might be looking at 50% or more non-image data. A file might contain a prologue or thumbnail data in binary format.

The software had generated a large chunk of data which were `private' to the software and its handling of that particular image. A file might also have trailing non-image data after the end-of-file marker `%%EOF' corrupting the final postscript produced on the archive. All these non-image data could be directly deleted from the file but the operator might also be looking at inefficiently trimmed image coordinates and definitions of fonts and colour schemes that were not actually used in the image.

The status of these figures had to be communicated somehow and with some suggestions on what to do. To send a message saying that "the figures have totally expendable Adobe Illustrator PrivateData" or that "simply removing everything after %%EOF cures the problem", might not be all that useful if such suggestions did not convey precise task-specific information. For someone who was not used to reading images as text, it was not immediately clear how to edit an image or even what editor can be safely used. The operators also had to be able to make sense of the incoming messages on how the software behaved.

Fragment 7:2

```
1 […] I'm 100 MB [Megabytes] over. The problem is that Adobe Illustrator
2 makes big .eps files and on my Mac where I do everything for papers
3 there's no way to make it smaller. […]
```

To produce figures that make the submission package a full "100 [megabytes] over [the size limit and] […] there's no way to make it smaller" on a Macintosh platform was simply not credible information. But it was taken seriously at the same time that submitters persistently claimed lack of control over the production of these Illustrator and Photoshop generated EPS figures. They claimed to have no idea why or how the software was operating as it did. One operator reported locally to the others about:

Fragment 7:3

```
1 […] flipping through books on Adobe Illustrator and all the claims by
2 various authors that precision (decimals in coordinates, resolution,
3 depth, you name it) cannot easily be tuned down to reasonable values are
4 bogus. I don't fall for that any longer, authors need to read the manuals
5 for their graphics software.
```

Suggesting the manual was one way of throwing the problem and the responsibility for it back into the hands of submitters, although with mixed results. Apparently these programmes were configured by default to include non-image specific data whether or not the user intended for that to happen. One of the things that transpired was that operators were saving time and efforts by cutting manually the excess data from these files. Then they inserted the packages into the pending queue and sent notices to respective submitters, for example,

Fragment 7:4

```
1 [w]e like to point out that all your figures were insufficiently
2 generated. They all included non-image data; thumbnails and unnecessary
3 color information. We processed 4 of them on our end.  Please make sure
4 you read the manual for  Adobe Illustrator(R) 9.0 were the instructions
5 are for generating minimun size postscripts excluding all non-image data,
6 extra fonts,thumbnails and including only the minimum color information.
7 […]
```

For the operators, these scenarios were rather puzzling. They openly enforced a size limit and were trained to use basic algorithms to manipulate figures and efficiently compress them. Now they were seeing reduction in file size to the factor of up to 10. But sending out messages claiming, for example, that `“[w]e know that the manual for Illustrator tells you how to generate postscript in the most efficient manner”` was hardly practical information. The operators could not be sure what these programmes were actually doing and, therefore, could not reference a manual with precise directives unless they had `Illustrator` or `Photoshop` up and running in front of them. And, since they could not communicate precise intelligible information, it was unclear how to proceed. It was not unproblematic either to make sense of lengthy explanations from submitters who were committed to making themselves useful. Consider this message.

Fragment 7:5

```
1 […] The way I obtained reasonably-sized files was the following.  Working
2 on a Macintosh, I opened each jpg file with Adobe Illustrator, and using
3 the Print function, wrote each to a file, choosing pdf format with jpeg
4 compression and no downscaling.  Then I opened each pdf plot in Adobe
5 Acrobat, and, again using the Print function, wrote each to a file using
6 eps with no preview.  For each file, the final version was only about 40Kb
7 larger than the original jpg.  (By the way, this was not explained in the
8 Illustrator manual.  If I "print" a plot directly from Illustrator to a
9 postscript file, it still ends up half the size I get by saving it as
```

10 Illustrator eps, which is 10 times the size of the original jpg.) […]

The operators could only guess what was going on from this message and the fact that: (1) functions like `Print' and `Save As' are universally implemented in software to create and store documents in a range of formats[10]; and (2) options to compress figures are widely implemented in imaging software. This submitter is using Illustrator to access JPEG figures (line 2), chooses a `Print' function to export to a PDF file, and names a menu option for JPEG compression and scaling of the image in the PDF format (line 3-4). Thereafter the submitter claims to choose a `Print' function from a menu in Adobe Acrobat, where the PDF is now displayed, and points out a menu option there for generating thumbnail data, or not, in the process of producing a postscript, "using eps with no preview" (lines 5-6). The outcome is 40 kilobytes of additional figure data compared to the original, which the submitter claims is "reasonably-sized" (line 1).

These processes cannot be immediately reproduced on the back end but the message is instructive. What is written into parenthesis in this message is also of some consequence. It was "not explained in the Illustrator manual" that using the `Print' function to create a postscript from the original gives "half the size I get by saving it as Illustrator eps". The operators could not *do* much with this information however, ie., produce credible instructions and help pages. They had no access to the particular ways in which figures could be produced with software they were not familiar with. They needed the material resources to reproduce the processes, experiment and unravel some of the contingencies locally. Accordingly, the reasonable thing to do was to install Illustrator and Photoshop on a `windows box'.

In this and the previous section, I have consulted and reconstructed a small selection of scenarios presenting challenges we did not see in the previous two chapters. We previously learned that the operators expected to see disorderly preprint files and bibliographic data.

[10] A `print' function essentially exports file data either to be interpreted by printers or to be `printed' to file in some format.

They inspected records, reports and files for anomalies. They made judgements about corrective measures or removal which is manifested in communication with submitters and computing supports and exchanges between the operators. In this chapter however, submitters and operators are exchanging information on corrective measures which do not refer to more or less familiar, task-specific and ready-to-hand phenomena. Nevertheless, problems with preprints were eventually resolved through trial and error.

In the final sections, I will relate these goings-on to issues of usability and access to the circumstances of doing technically-informed work in the daily administration of preprints. Next, I will consult and reconstruct communication which took place between the operators, new members on board the archive, and senior managers.

Planning `talk'

On my list of things to do during my time in the daily administration of preprints was to analyse and assess the application software used to do this work. The idea was to get a better understanding of how staff at the library could contribute to this work in the future and what alterations would be needed. Originally my colleague and I had been asked, rather casually, to get involved with archive operations, which we did. Nothing was officially planned about this involvement until a meeting was called a few months after our involvement first began. In the meantime, I had become one of the operators in the daily grind, while my colleague had been observing the message box and help documents but had not actually done any of the administrative work.

Everyone at the university who had stake in this work was present at the meeting: (1) operators who operated the archive in its previous location, one of them the inventor of archive; (2) my colleague I (programmer/analysts at the library); (3) the Dean of the computer science department (the inventor's superior); (4) the university librarian with two

associates, one being the director of the division of digital library and information technologies (my supervisor), and the other the director of the physical sciences library.

My colleague, the operators, and I had been brought together to this meeting with senior managers to discuss the administration of preprints and how to move forward. The work was described by my colleague as rather `mindless' manual routine labour. We both suggested that quality checking bibliographic data, checking document files and removing unwanted or corrupted materials could be separated from more expert-intensive tasks. What my colleague and I had learned was how much of this work required relatively simple but technically-informed judgements which then could only be acted on in expert-intensive ways. The inventor argued that the apparatus was sufficiently configured to do an enormous amount of inspection, detection and auto-correction with respect to the bibliographic data. Then another operator proceeded to summarise and clarify, in general terms, how the finer and courser grains of inspection, detection and judgement were delegated between system components, submitters and operators. We discussed this back and forth and my interest in these matters, as I expressed it then and shared with my library colleague, was to establish an agreement at this meeting that the administrative environment could be simplified significantly and what amounted to clerical support should be redesigned in a more user-friendly manner. The other operators did not object to the idea at all, nor to our suggestion that the syntax of tools and supports to do clerical work could be in a different language and the environment graphically interfaced.

According to my meeting notes, we came to an agreement that a redesign of interfaces should result in simpler administrative set-up where tasks could be distinguished in terms of *technical* and *non-technical* personnel. My colleague and I were asked to analyse the current arrangement and, to do so, my hands-on experience, which I was acquiring at the time, should come in handy. But we also agreed that it was foreseeable that l*ow* and *high* skill

labour, as we put it, would remain entangled for some time to come. We agreed to experiment and figure out how to coordinate efforts on an ongoing basis.

It was never addressed at this meeting what it actually meant to say *technical* and *non-technical, low* and *high* skill personnel, ie., what exactly they were supposed to be capable of. Neither did we address any technical specifics of how the administrative environment was constituted and worked on a daily basis. This was neither the time nor the place. The senior personnel sitting toward the right end of the meeting table were mostly listening in on our discussion. They asked questions and casually gestured approval, suggesting that we knew what we were talking about over on the left end, and thus a good reason to support our ideas, namely, that my colleague and I would work closely with the experienced operators to produce an assessment document.

This meeting could be described as a committee meeting in which seniors are briefed and give their blessings to suggestions from their underlings on how to proceed. Our meeting tools were our agendas: questions and arguments we wanted to raise *about* the apparatus and what to plan for it next. The discussion also moved on to address issues regarding submitter status and other policy-related matters of interest to the science librarian, available funds, and so on. But assessment and redesign, the topic I have drawn attention to here, had now been *formally* put into motion.

With hindsight, there were uncertainties after this meeting about how to proceed and, as a matter of curiosity, it took more than half a decade to accomplish precisely what was suggested in this meeting—ie., a decoupling of clerical work from expert-intensive support. There had already been confusion surrounding the set-up of the new main server machine at the university—who should get involved from the library and what precisely their role should be. There was also confusion surrounding first attempts by library personnel to engage in the daily administration of preprints, as I will now attend to.

More planning `talk'

A letter dated on 5th September 2001 was circulated to propose topics for a meeting. Four staff members at the library were recipients of the letter and thus involved in first talks about, as it was phrased, "working to begin the transition of the day-to-day operation". It is stated in this message that the group wants to,

Fragment 7:6

```
1 know the specific tasks that get done daily and would also like to have
2 a discussion about how we can begin to answer some of the [incoming
3 mail] questions. Another, more general, topic for the meeting--probably
4 to be continued later--might be the requirements for a set of scripts
5 to automate some of the routine administrative procedures.
```

My colleague, the author of this letter, had been looking for a while at the "questions", ie., the message box, but was not set up to handle matters that needed attention. There had not been any direct hands-on involvement by a library employee at this point in time. This message mentions having "a discussion about how we can begin to answer some of the [incoming mail] questions" (lines 2-3). According to my notes, however, the operators were already waiting at the time for someone to `enter' the environment, orient themselves and figure out how things worked. An earlier exchange on the matter indicates also that, among library personnel, their orientation was to issues of delegation and the future organisation of the work rather than working the apparatus as it was. This is what they wanted to discuss.

Fragment 7:7

```
1 Sometime soon we will have to figure out a workflow for contributing to
2 or taking over the administrative mailing list, too. I've been lurking
3 the list and have been studying the help documentation, so I have a
4 beginner's knowledge of what kind of questions and answers we will be
5 responsible for, but I'm not ready for the transfer yet.
```

In these two excerpts (fragm. 7:6 and 7:7), there are no remarks to the effect that learning hands-on to cope with the apparatus should be the first step for an operator to be "ready for

the transfer" (fragm. 7:7, line 5). My colleague mentions the need for efforts to "figure out a workflow for contributing or taking over the […] mailing list" (lines 1-2), not mentioning the possibility of a perfectly useful pre-existing workflow to which a new operator should get acquainted as quickly as possible. The author mentions "studying the help documentation" and having some "knowledge of what kind of questions and answers we will be responsible for" (lines 3-5). There is no mention of *doing* the work, only that the author is "not ready for the transfer".

It is also mentioned in fragment 7:6 that the administrative work needs automated procedures. The operators had mentioned that getting on board the daily grind should involve new software development. However, the repeated requests to meet to communicate what tasks to transfer and how, as opposed to getting `hands dirty' and *doing* the work, left members on both ends somewhat confused about how to proceed. This message in fragment 7:6 contains a request for a meeting agenda to discuss the "requirements for a set of scripts" (lines 4-5) in reference to automation and routine administrative procedures. But, according to what we have learned so far, a range of automated processing took care of routine inspection and diagnosis, and there were a number of scripts ready-to-hand, although the operators freely admitted that these instruments could very well be improved or programmed and arranged differently. I will now consult further exchange between archive operators (**AA** and **AB**) and my colleague (**LS**). A third operator is mentioned (**AC**).

On Monday 24th of September 2001, **AA** sends out a message, addressed to **LS**:

Fragment 7:8

```
1 AC is doing the daily grind this week and I'm trying to sort out
2 remaining changes needed on the mirrors before we switch the main site
3 […].
4
5 I suggest that to get a real feel for the daily grind, you might join
6 me one day next week from 2pm to a little after 4pm […].
```

According to this excerpt, **AA** is busy with development and maintenance work and suggests to **LS** (and me) to join up sometime in the week thereafter to get a "real feel for the

daily grind" during the afternoon peak time (line 5-6). We would see for ourselves the work as it was typically performed at an operator's desktop server. On the day after, Tuesday the 25th, an exchange between **AA** and **AB** is copied to us.

Fragment 7:9

```
 1 It was my understanding that no labor is in place yet anyway and I
 2 really do need to get the mirrors sorted out this week.
 3 --
 4 AA
 5
 6 AB writes [quoting message from AA]:
 7 > > I suggest that to get a real feel for the daily grind, you might
 8 > > join me one day next week from 2pm to a little after 4pm […].
 9 >
10 > except do we really have to wait that long to start getting a labor
11 > complement in place?  wouldn't it be best to start training a triage
12 > person a.s.a.p.?
```

Lines 10-12 are **AB**'s remarks in response to **AA**'s message to **LS** on the previous day (lines 7-8). **AB** expresses concern about the session not taking place until "one day next week", as **AA** had suggested (line 8), and asks if "a triage person" should not be trained as soon as possible. The daily administration needed someone on board to assess, inspect, prioritise, diagnose and, accordingly, take further action—in a word, a triage person to oversee an average of 140 preprints awaiting public dissemination. **AA** is clarifying his position, in response to the remarks by **AB**. He is busy with pressing matters and does not think that "labor is in place yet anyway" (lines 1-2). As I found out, **AA** had been notified earlier about me joining **LS** for this visit but **AA** does not seem to be expecting labour and, according to this message, is not in a rush to meet us. On the 26th, **LS** sends the following note in response to this exchange:

Fragment 7:10

```
1 AA,
2 Perhaps I didn't communicate this clearly--Kristrun Gunnarsdottir
3 and I _are_ the labor until we hire some non-programmer staff to
4 take over. I believe we'll be able to judge how best to structure
5 support in the library after we do some of it ourselves.
```

In the proposed visit to **AA**'s office, **AA** was going to show us and demonstrate the daily

grind. We would see, for the first time, an operator do what needed doing and have opportunities to ask questions. But, the confusion about how exactly to proceed with this collaboration is manifested in the naming of `labour'. First, there was a complaint. The operator was waiting for "a labor complement" (fragm. 7:9, lines 10-11), followed by a response from another operator that there was "no labor in place yet" (fragm. 7:9, line 1) followed again with an explanation from **LS** that we "_are_ the labor" (fragm. 7:10, line 3). But **LS** also mentions "some non-programmer staff" (line 3) as opposed to *programmer staff*. I soon became aware that the "labor in place" or "labor complement in place" was not the labour **LS** was offering on behalf of the library. **LS** wanted to be able to "judge how best to structure support in the library" so that "some non-programmer staff" could be hired to do a then still undefined job. There is no mention in this letter, or in other exchanges around that time, what programmer staff or other competent technical skill was available as operators. The statement in this letter is that `we' are programmer staff, the labour charged with the task to structure support for the daily grind and therefore we should be doing "some of it [the administrative work] ourselves". (line 5). In other words, our involvement was meant to be partial and a temporary arrangement for other purposes than actually working in the daily grind.

Working the `daily grind'

In this section I will again draw attention to the situation facing the operator at the interface with two excerpts from my recording of explaining this work while doing it. Both concern the task of removing a preprint and what is recorded are comments on what is taking place. As I explained in Chapter 3, this recording was an attempt to recover as much as possible of doing the administrative work and then present the transcript to other staff members at the library for analysis and further considerations [see also Appendix IV].

Fragment 7:11

```
 1 I should move quickly through the PDFs because it looks like I am going to run
 2 out of time, it is 15:43. We can have the scans through PDFs [checking through
 3 a list of them], which is an easy job, somehow flagged to office personnel…
 4 (.)  they may have to be removed. What is problematic about them is that the
 5 removal process at this point is not very straightforward for a GUI
 6 [graphical/web user interface, (also WUI)] oriented computer user.  (.)  We
 7 could see that better automated with GUI…  (.)  button stuff.  (.)
 8 […]
 9
10 Here I am checking for moderator comments in the busybody mail just in case and
11 this is important to do before [the] 16:00 [freeze]. Here is a suggestion that
12 a paper submitted to `nucl-th' be submitted to `cond-mat' and since we always
13 do what they say I will have to route. Now, what I really do is that I
14 basically remove the paper, as previously described [with `lastmv'], jamm, no
15 kidding, and then I move myself into a sub-directory `TMP' and `reput' the
16 paper into the system from there, pipe it into the other news group…  (.)
17  `cond-mat'…
18 […]
```

These comments are instructive. They are spur-of-the-moment thoughts on what is going on. For example, on lines 2-3 the comment is made that `"[w]e can have the scans through PDFs [checking through a list of them], which is an easy job, somehow flagged to office personnel…"`. But this remark is also followed by, `"[w]hat is problematic about them is that the removal process at this point is not very straightforward for a GUI [or WUI] oriented computer user"` (lines 4-7). The routing of preprints between subject archives is also commented on by explaining that `"what I really do is that I basically remove the paper, as previously described [with `lastmv' …], and then I move myself into a sub-directory `TMP' and `reput' the paper into the system".`

Both `lastmv` and `reput` were command-line devices invoked with specific parameters to direct their operations. To remove, the **`archive/papernum`** was needed and to manually `pipe' a package through, the `reput` script needed the temporary identifier and the name of the relative subject archive, `cond-mat`, `nucl-th`, and so on. One could argue that the devices used to remove from the system, `lastmv` or `lastrm`, were easy enough to explain to those who came new on board. They could be invoked without knowing what they were actually doing in the background other than the fact that they removed all traces of a preprint in the pending queue. However, their uses had contingencies. They would `talk' back in a hybrid

language and ask for confirmations. We learned in Chapter 6 that *not* using either of them, but using `rm' instead, resulted in `SERIOUS ERROR'. But then again, there were occasionally situations in which using `rm' might be relevant, also the generic `mv' command to move one file over another, ie., in the case of removing a new version to an existing paper in the system by `rolling back' to the previous one.

Having been an operator in the daily grind, I can verify that there was nothing particularly unreasonable about these arrangements and procedures. Nor was it unreasonable to thread communication in the message box in order to separate tasks and thereby create `holding areas' for unresolved issues and make choices about what to do first and what could wait. It made sense to use a local desktop environment to copy things over for diagnosis and repair to avoid accidental damage to packages on the server machine. Workarounds to save time and efforts (eg. if a `PS BAD' warning, check the PDF first) made perfect sense as well, so on and so forth. Furthermore, there was nothing unintelligible about the fact that the daily administration of preprints centred on inspection, the detection of anomalies and procedures to act on problems that occurred. That was the very reason for doing a job of being an operator in the daily grind. The devices to do that work however, and act on judgement calls, increasingly became a puzzle with respect to any kind of redesign considerations.

What these efforts to capture the work at the interface achieved at the time was to show, or rather *translate,* to colleagues at the library the potential confusions in getting seemingly simple tasks done. The work on that day when this recording was made involved, for example, `piping' a registration through manually, which was needed if auto-registration failed because the request came from a facility which did not clear automatically but the registration request was nevertheless acceptable. Tasks that were well recognised in libraries were reported on this day, such as quality control of bibliographic data, concern with copyright and action taken in breach of it. Then there were less familiar tasks such as the expert-intensive procedures of orienting and attending to some of the problems submitters

communicated: problems associated with using the TeX auto-compiler or corrupted images. How to remove or `roll-back' a package is also a case in point. My discussion of a mistake in command in the previous chapter, unravels many interlinking parts and data relational dependencies, "password entries, title databases, counters, etc. to reset, too". They were wide open to breach. The scenario I consulted in Chapter 5, fragment 5:15, is also a case in point. We see what is communicated about pscheck, what it does and why it reacted in the way it did. One could argue that it was easy enough to explain what the relations were between certain functions of hack and pscheck, what pscheck precisely does and when it should be invoked. What we see however in fragment 5:15, is how very basic relationships which *should be obvious* to an operator are missed. In sum, the one thing that rendered clear was how coping with the daily grind required advanced computer literacy because tasks were generally carried out in expert-intensive ways and tools and supports encoded in languages like PERL, TeX and the Linux assembly language. The devices to do this work were not usable to most office workers in a library.

Discussion

In this chapter, I have done much less in the way of showing the reader the details of doing work, but draw more on my personal and professional involvement and use my own accounts along with recorded communication on the archive to support an argument I will now make about the relevance of attending to such details. In the first scenario (fragm. 7:1), we see how an operator offers nothing to a user of Scientific Workplace, only what is relevant to the use of TeX. The onus is on the submitter to interact with the phenomena the operator is attending to and the ways in which the operator orients the submitter to them. The matter is left where the exchange ends and the preprint was submitted. However, as I pointed out in my discussion of this sequence, these attempts were, at best, fragmented efforts to establish alignment and move forward. The submitter does not know how to use text

encoding or where to insert the relevant command. The operator remarks that "the sciword tech support should […] know" and, according to the exchange we see between them, sciword tech support should attend to problems in using the product, rather than an operator on the archive, ie., those who sold the submitter software which allegedly exports to standard LaTeX.

In a series of communications about figures produced with Adobe imaging tools, we are made aware that submitters claim not to understand what the software is doing and the operators were somewhat puzzled by what they saw in these figures, ie., 50% or more non-image data (fragm. 7:2-7:5). This problem with bloated file size was not entirely ignored. It kept coming back. Submitters offered advice on how to compress figures more efficiently (fragm. 7:5). An operator claimed to have flipped "through books on Adobe Illustrator" (fragm. 7:3, line1) and thinks authors are making bogus claims and should be consulting manuals properly. Incredible claims were made, eg., "there's no way to make it smaller", a figure which is 100 megabytes over the submission limit. What we learn is that the operators could not adjust their practices, provide credible advice or implement automated procedures to detect and remove the excess data, without investigating the very circumstances in which these figures were generated in the first place, the details of which can only be known by *doing that work*.

When I turn the attention to events taking place at the library, we learn that expectations differ between operators and library staff with respect to the question of what needs doing to be adequately involved. The unfolding events have resonance in the `classic' story told by ethnomethodologists and other workplace ethnographers (eg. Button and Harper, 1998; Heath and Luff, 2000, p.1-8; see also Chapter 3 and discussions in Button and Dourish, 1996; Suchman, 1983; Suchman,1998; Trigg et al., 1999). It tells a story of the computerisation of a workplace in which people already go about their business in a more or less orderly fashion: index cards and notes are scattered about and things are shuffled around relative to the ebb

and flow of the work. When designers and engineers come onto the scene, they do not pay adequate attention to the details of this work and design work systems that complicate procedures or completely fail. In this `classic' story, the workers are also often blamed for not appreciating the benefits of new systems. The daily administration of preprints on the archive was a peculiar variant of this story. It was already computerised, but the supports were not a purpose-built comprehensive support system. Rather, as we learned in the previous two chapters, worksite supports were ad hoc reactions to the fact that submissions and submitters where not always reliable. Implementing auto-inspection and diagnosis, defining specific warnings and assembling tools to deal with matters, could only be realised on the basis of local knowledge and experience at any given time. What was unique about the daily grind was that those who worked it were all in one, experts in computing, authors of the apparatus and the workers who did the work with intimate knowledge of all its goings-on. First attempts by library personnel to get involved were more or less of the `species' which is addressed and critiqued in studies of workplaces and computer supports. They were at first not prepared to put in the work that would enable adequate attention to the details of how precisely matters were arranged and the work accomplished. But the operators knew better and demanded more involvement.

In the final section on working the `daily grind', I offer another account of doing this work at the interface with some spur-of-the-moment comments. In my discussion of fragment 7:11, I relate its contents to issues of usability and the usefulness of tools and procedures. It was never pre-given what exact skill to look for in new operators or assistants apart from being familiar with `Linux` and `Perl`, and ready to learn `TeX`. It became an empirical matter to integrate new staff. But it was very instructive to actually do this work. It became increasingly more difficult to imaginatively simplify procedures of inspection and attending to anomalies. However, the practical relevance gradually emerged of *seeing for oneself* the phenomena that needed attention and how they were attended to. Briefly, the decoupling of

clerical work and expert-intensive support did not happen while I was on board, although plans were devised at the `committee' meeting referred to in this chapter. There was no radical disruption of existing arrangements and procedures. But a level of partitioning was achieved with assistant techies (engineering students) and expert-intensive support dividing the labour. It also transpired that the underlying structures of the apparatus needed major overhaul and the operators articulated what needed attention. Steps were taken to re-constitute structural components of the system, however, attending to unreliable submitters/submissions had to take its natural course for the time being.

The events involving bloated figures and developments in collaboration with library staff draw attention to the practical relevance of adequate attention to detail. What transpires is that, in and through intimate acquaintance with existing arrangements, tools and procedures, a site—on which that effort is made—becomes *the* platform for practical and meaningful developments. New contributions—ways of working, object handling, etc.—were both absorbed into the processing environment and constrained by already existing configurations. These relationships were not fixed but fluid and contestable, and things did change over time.

Chapter 8

Concluding remarks

A summary of the thesis

I will now summarise my findings and revisit some of the discussions I developed throughout this thesis. As I explained in the introductory chapter, I was primarily concerned with the everyday management of complex virtual objects – the uploading and downloading of scientific articles – and the interactions of operators and submitters to ensure a robust and consistent service. A central concern in this thesis has been to respecify problems of disorder in technical work, which I took as my point of departure, to enable the reader to see phenomena of on-going detective, reactionary and preventive labour.

In Chapter 2, I clarified how studies in ethnomethodology stand apart from methods and reasoning in conventional sociology. My concern with the matter was provoked, in part, by the material I set out to work with. I had access to details of the technical work that needed doing. I had first-hand knowledge of the circumstances of actually doing that work. Consequently, I was never comfortable to assume an idealised *standpoint* from where I could *look in*, define, characterise and assess what was going on from a *perspective* of sociological reasoning. I was not willing to commitment myself to explanations involving social actors, social forces, structural and normative constraints using the language of causal relations and causal adequacy. Studies in ethnomethodology have been a resource for me to see in the material at hand the methods that actors use to do what they do and to show their commitment to matters of origin, circumstance, practical responsibility, competence, legitimacy and cooperation. But rather than adhering to the logics of structure-agent, actor-system dichotomies, the people who are involved in the operation and use of the archive are always already, to paraphrase Garfinkel, in their local, endogenous, natural organisation,

which is everywhere entirely their own work. From their point of view, some form or another of order, knowledge, skill, etc., is always already relevant to the purposes at hand.

By suggesting that we should *see* the work that took place in the operation and use of the archive, I also set out to show the reader how tools and supports were committed to do their share of this work. It is not contested that computerised document and record systems are, by their very configuration, ideal to *sort things out*. They are constructed to organise object classes and relations in a consistent and reliable fashion. But they are also equipped with interface supports that organise labour through predefined and restrictive procedures – eg. data input, search, analysis, delete, download, etc. As I discussed in Chapter 3, studies of worksites have made a strong case of criticising common ways of approaching the design and engineering of computer supports and workplace systems. This critique has been aimed largely at cognitivist assumptions and formal descriptions that commonly underpin designs. How people actually organise and do their work is either overlooked or ignored and the consequence is that tools and supports are designed in ways that are not amenable to naturally occurring contingencies that arise in practice.

The situation with the archive was somewhat different. The archive apparatus could automatically upload and download articles according to guidelines and with efforts to solve any issues at the front end. The work that took place in the daily administration of preprints did not centre on those objects. Rather, the work that needed doing was organised on the assumption that preprints and bibliographic data could always be *out of order*. But the tools and supports to inspect submissions and sort out problems were not purpose-built. They were built piecemeal in response to what had been learned about submitters and their submissions over time, and what could be done to manage preprints and bibliographic data in spite of the evident lack of uniformity. It begs the question of what we expect of computers in coping with disorders.

To enable the reader to see what is said and done on the archive, I set out to capture and reconstruct occurrences in ways that should be recognisable and transparent. I explained in

Chapter 4 that unified methodological protocols and descriptive models are not part of the programme of ethnomethodology. However, the `unique adequacy' requirement, as I understand it, concerns a method for demonstrating what a description can say about a practice by capturing the complex surface textures of what goes on. But methods have also been treated as discoverable in the course of my involvement with archive and in writing this thesis. As I explained in Chapter 4, `data' was collected as a matter of course. I have relied on my ability to adequately describe technical arrangements and reconstruct tasks by drawing on electronic records and my long-term involvement. Distinct topics of disorder is what I found in the details of lived action and interaction and, as the reader will have learned, I have assumed there are more or less orderly, mundane and business-as-usual practices of coping with the unusual and the unexpected. I treated my own work on the archive as a relevant object of enquiry by using recordings of doing the administrative work as an employee, but they were produced for local practical purposes. I have also taken liberties to draw on my experience to describe events, developments and assessments to the best of my ability to tell a coherent story, and I have communicated with former colleagues about my study interests and the contents of this thesis. The topics that my methods gave rise to were ordered into Chapters, 5, 6 and 7 which can now be summarised.

Chapter 5 focused on the work of inspecting submissions when they were uploaded on the archive. The main sections drew attention to that which, on inspection, was incorrect, incomplete, corrupted, irresponsible or otherwise illegitimate. By consulting a number of scenarios in doing this work, we saw that the operators commit to questions of completeness, origin, constraints and practical responsibility. Submissions/submitters could be unreliable and inconsistent and we find that there were obvious limitations to the contributions of computer supports in detecting the anomalies. What we see is how socially relevant matters, to which operators were committed, attach themselves to specific phenomena that *can* be seen but typically also demand further inspection. Operators rely on computer supports to mediate data traffic, auto-diagnosis, histories and communication. For example, the apparatus

had been committed to questions of legitimacy by flagging, for human inspection, all 2-page submissions, all exclamation marks in abstracts, all single-file `PDFs` and `postscripts`, and more. Only the eye could catch subtle mismatches and anomalies. Auto-generated warnings were flagged and found and what we learned is that there are no hard and fast rules to this game. Seeing how this work is done reveals a complex relationship between computational functions and subtle human judgement.

Chapter 6 focused on procedures of taking corrective actions and communicating problems and solutions. The main sections drew attention to scenarios that required direct communication between operators and submitters, and direct involvement on the part of operators. I foregrounded how `making efforts' was significant to the ways in which operators and submitters cooperated and coped with each other and the archive apparatus. What we see in Chapter 6, which corresponds with Chapter 5, is how *looking for* and *seeing* takes precedence in getting on with the work that needs doing. Submitters and operators name the tools they use, the phenomena they see and want others to see, and they orient themselves and others to exact locations of where one should look. We learned how particular the submitters and operators could be about encoding the details of an act or a command. For example, we see how a submitter orients operator's attention away from what can be eliminated and towards the location of a problem as seen by that submitter. We see how an operator orients a submitter to evidence that not enough effort has been made and how they exchange information on tools and procedures—exchange evidence of efforts—to establish how to move forward. We learned how a competent and accountable use of the tools was established in and through these interactions and exchanges.

Chapter 7 took examples of scenarios in which submitters and operators were challenged by unfamiliar tools and procedures. Submitters were having problems with preprints prepared on `Windows` and `Macintosh` platforms. This presented problems in the exchange of information between submitters and operators. They were not using the *same* tools and thus not able to refer one another to more or less familiar task-specific instructions. The apparatus

was also settling down in a new location where library employees became involved in the daily administration of preprints. The possibility of establishing a collaboration between library personnel and the archive operators led to preliminary plans to assess how the library might contribute to this work. As I explained in Chapter 7, the events that unfolded resonate in a peculiar way with the critiques I discussed in Chapter 3. The classic story is one of computerising workplaces, where systems are put into place without adequate regard for existing practices and the lived details of doing the work. Accordingly, they either complicate or fail to do what was intended with the design. The daily administration on the archive was computerised, although, as I have explained, tools and supports were not purpose-built but implemented to meet the practical needs of doing the work as such needs emerged. This had evolved in the hands of those who were both experts in computing and, as operators, those who worked the system. In this story, programmers/analysts at the library suggested a restructuring of the daily grind but, to do so, they had no choice but to commit to elaborate involvement with the particular ways in which this work was already achieved. I will now move on to consolidate some of the discussions I have already provided and reflect on the implications.

Completeness and consistency

Programming is certainly one central task of integrating designs into ways of getting work done. Although design issues are not central to this story, they were prominent in my involvement in the archive. What could be realistically delegated to devices? What can be implemented without disrupting already embodied and successful work procedures (see Nichols and Twidale, 1999)? Knowing instrumental rationality is essential to this line of enquiry. But so is keeping a check on the limits of resources available to computational constituents in relation to what is expected of them in the workplace. The *pure dissemination system* promoted by the founder of the archive only had to ingest and disseminate data packages without a halt. In this very narrow sense the system was indeed automated because,

as far as the auto-processing devices were concerned, leaving the throughput exclusively to their crunching would be unproblematic. However, the extent to which it is possible to sustain such an arrangement, or the argument that it is a *pure* system, raises the question I addressed in Chapter 4 about error levels. In document and records management, there is really no level of error acceptable. One of the operators suggested to me that it was a matter of policy whether an automated function undermines what you are trying to accomplish or improves it. According to my observations, it is–as practitioners themselves typically see it with or without an explicit policy–a matter of completeness and consistency. This applies to library services, record keeping and management in law offices, accounting offices, and the like. The same applied to the archive as a growing storehouse of scholarly works and a service to users. What was at stake was the completeness and consistency of a service to everyday practices amongst scientists and mathematicians. That said, what was interesting about design and engineering decisions, built into the daily grind over time, was how deviations and anomalies were responded to by implementing harvesting devices and probes rather than an attempt to *fully* automate. Operators developed multiple layering of auto-processing, auto-detection, auto-correction and *flagging*, with varying degrees of inspection and manual interception. What I also witnessed on site was that when new anomalies and deviations emerged, the operators would communicate if they could be *better flagged* or how to auto-detect them for intervention. They would not implement auto-corrections unless they were shown to be either strictly reliable or could otherwise be implemented with a prompt for confirmation. In other words, it was no mystery to the operators what could or could not be delegated to machines and how delegation could safely support the kind of work they were doing, namely, by preserving flexibility in the relationship between particular computational functions and subtle human judgements. There was no evidence of oversimplifications or conditions halting the access to the apparatus in some way and making it impossible to respond satisfactorily to a judgement call. There was no evidence of encoded aims, motives or commitment to particular actions. There was, in effect, less programming rather than more

of it. This finding supports the arguments raised by Suchman (1998), Button and Harper (1996), and many other scholars who have concerned themselves with struggles in workplaces which are brought about by conflicts between design implementations and particular needs emerging in the course of getting work done. What this case has to offer to that discourse is that the operators, as experts in computing, remained in close proximity to the practical reasoning in actually facilitating consistency and completeness. The result was an enormous assembly of building blocks, tool boxes, and open-ended futures.

Problems of disorder

Ideal-type engineering can be defined by the absence of deviation and anomalies. Classification schemes and categories are structured into comprehensive systems into which uniform auto-processing can be introduced with only minimal interception to steer it forward. Such engineering uses traditional formal logic with disregard to all the fringes. Manual labour should be minimised or eliminated, standardisation maximised, purpose-built devices put into production to take over and support specialised tasks, so on and so forth. The design and engineering practice itself, if it was in effect ideal, concerns a world in which the concrete actualities of people at work, grappling to produce a coherent ordering of objects and environments to manage them, for all practical purposes, are lost. It becomes a *theological* idea (Lynch, 1993, p.151, cf. Derrida, 1970, p.256). But as Button and Sharrock (1998) point out, engineers are always working with less-than-ideal circumstances. They are, I argue, working with naturally occurring disorders of the world.

In the case of the archive, we have seen that submissions will fail or require attention because they do not conform to guidelines or material constrictions. It is a taken-for-granted assumption that there are submitters who follow guidelines the best they can and there are others that make a poor work of it or do not read them at all. There are submitters who make their best efforts and others who only do the bare minimum. There are those who claim to investigate and have taken certain actions which sometimes are incredible claims. There are

submitters who will draw attention to themselves, blatantly break rules, post badly formatted bibliographic data, inappropriate or incomplete preprints. There are also submissions that are on the fringes of what the apparatus can accommodate but, it is *always possible* to integrate new variations. To remind ourselves, most submissions were taken care of on the front end and most submitters did not communicate with operators but, as we also established, the work of operating the daily grind constituted a function, a computer supported job. Are we looking at naturally occurring *orders* or *disorders*? I am not interested in the technical uses of these terms. The project here is not to try to understand why these differences occur either—to frame their existence as an aberration pitched against normative obligations to be responsible and orderly. Rather, my project is to open up these practices to inspection. Operators on the archive created measures to accommodate disorders in the sense that preprints are far from uniform in their material constitution and method to produce them but, nevertheless, they *can* be accommodated by a single apparatus. The operators also created measures to detect and harvest phenomena pointing to disorderly conduct, irresponsible origins and the like. The focus here has been to treat these activities as a case study which acknowledges that problems of disorder actually coincide with an urge, a *will to order*. The many phenomena of disorder gave shape to the computer supports, ongoing modification of procedures, new methods of flagging warnings, developing auto-detection and harvesting protocols, and more.

The issue I raised in the opening chapter drew on Jordan and Lynch (1992), that conditions of instability and fragmentation in routine technical work are obvious to the technicians themselves as well as to those who use their products. But, the preoccupation of ethnomethodologists has been with the ways in which orders are achieved and how business is normally conducted. This preoccupation with the ways in which orders are achieved teaches us how people ordinarily `put things in common', make sense of what they are working with, make sense of each other and *sort things out*. Thus the *orders* achieved attach themselves to the ways in which phenomena are reportably-accountable and mutually

intelligible in manifesting practical issues at stake in a coherent manner. But this focus of enquiry leaves us with very little done in the way of treating pre-theoretical phenomena of disorders as phenomena of considerable interest to be studied in the own right.

The core of the work being done here only gives a glimpse of what the phenomena were and how they were dealt with in the daily administration of preprints on the archive. My task has been to show the reader and explain some of the phenomena manifesting that which, on inspection, was deemed incorrect, corrupted, incomplete, or not according to guidelines and material constraints. It indicates what *can* be uncovered if such phenomena were attended to adequately and the implications are wide ranging. For example, this thesis has implications for new and emerging technologies of surveillance and sorting: airport security, remote health monitoring, law enforcement, product tracking, and more. New systems are being designed and implemented specifically to detect actual or potential disorders or anomalies of conduct and risk—to flag and warn of dangers. This thesis already begs the question what we expect computer supports can contribute to inspection and detective work. What is expected of them to read, harvest, flag or auto-reject? What are they actually reading—what are the phenomena that can be known or detected and thus `anchor' culturally-specific and practically relevant formulations? How are they read, monitored and interacted with to verify that an anomaly, risk or danger is indeed what has been detected? These are timely questions to ask. For example, surveillance and monitoring technologies receive a lot of press and are thoroughly discussed in the circles of philosophers, ethicists and policy researchers to sort out potential issues of harm, breech of human rights and what might be the ethical, legal and social implications. What we know remarkably little about are their actual and potential uses on the ground, and the ways in which these new and emerging technologies actually serve the will to order.

Bibliography

Abbate, J. (1999). *Inventing the Internet*. Cambridge, MA: MIT Press.

Ackerman, M. S. and Halverson, C. A. (2004). Organizational Memory as Objects, Processes, and Trajectories: An Examination of Organizational Memory in Use. *Computer Supported Cooperative Work* **13**(2). pp. 155-89.

Adkinson, B. W. (1976). Federal government's support of information activities: A historical sketch. *Bulletin of the American Society for Information Science* **2**(8). pp. 24-6.

Agre, P. E. (2003). Information and Institutional Change: The Case of Digital Libraries. In A.P. Bishop, N.A. Van House and B.P. Buttenfield (eds) *Digital Library Use: Social Practice in Design and Evaluation*. Cambridge, MA: MIT Press. pp. 219-40.

Agre, P. E. (2000). Infrastructure and institutional change in the networked university. *Information, Communication & Society* **3**(4). pp. 494-507.

Agre, P. E. (1995). From High Tech to Human Tech: Empowerment, Measurement, and Social Studies of Computing. *Computer Supported Cooperative Work* **3**(2). pp. 167-95.

Akrich, M. (1992). The De-Scription of Technical Objects. In W.E. Bijker and J. Law (eds) *Shaping technology/building society : studies in sociotechnical change*. Cambridge, MA: MIT Press. pp. 205-24.

Allhoff, F., Lin, P., Moor, J., et al. (eds) (2007). *Nanoethics: the ethical and social implications of nanotechnology*. Wiley-Interscience.

Althusser, L. (1971). *Lenin and philosophy and other essays (translated by Ben Brewster)*. London: New Left Books.

Amoore, L. (2006). Biometric borders: Governing mobilities in the war on terror. *Political Geography* **25**(3). pp. 336-51.

Anderson, R. J. (1994). Representations and Requirements: The Value of Ethnography in System Design. *Human-Computer Interaction* **9**(2). pp. 151-82.

Aradau, C. and Van Munster, R. (2007). Governing Terrorism Through Risk: Taking Precautions, (un)Knowing the Future. *European Journal of International Relations* **13**(1). pp. 89-115.

Atkinson, P., Coffey, A., Delamont, S., et al. (eds) (2001). *Handbook of Ethnography*. London: Sage.

Bachrach, S. R., Berry, S., Blume, M., et al. (1998). Who Should "Own" Scientific Papers?. *Science* **281**(5382). pp. 1459-60. http://www.sciencemag.org/cgi/content/full/281/5382/1459

Barber, B. (1959). The Sociology of Science. In R.K. Merton, L. Broom and L.S. Cotterell Jr (eds) *Sociology Today: Problems and Prospects*. New York: Basic Books. pp. 215-28.

Barcellini, F., Détienne, F., Burkhardt, J., et al. (2005). Thematic coherence and quotation practices in OSS design-oriented online discussions. In M. Pendergast, K. Schmidt, G. Mark and M. Ackerman (eds) *Proceeding of the 2005 International ACM SIGGROUP Conference on Supporting Group Work*. Sanibel Island, Florida, USA. New York: ACM. pp. 177-86.

Barnes, B. (1974). *Scientific knowledge and sociological theory*. London: Routledge & Kegan Paul.

Barthelmess, P. and Anderson, K. M. (2002). A View of Software Development Environment Based on Activity Theory. *Computer Supported Cooperative Work* **11**(1-2). pp. 13-37.

Beck, U. (1992). *Risk Society: Towards a New Modernity (translated by M. Ritter from the German Risikogesellschaft (1986) w/introduction by S. Lash and B. Wynne)*. Sage Publications.

Becker, H. (1963). *Outsiders*. New York: The Free Press.

Benson, D. and Hughes, J. A. (1991). Method: evidence and inference - evidence and inference for ethnomethodology. In G. Button (ed) *Ethnomethodology and the human sciences*. Cambridge University Press. pp. 109-36.

Berg, M. (1998). The Politics of Technology: On Bringing Social Theory into Technological Design. Science. *Technology and Human Values* **23**(4). pp. 456-90.

Bernstein, A. (2000). How Can Cooperative Work Tools Support Dynamic Group Processes? Bridging the Specificity Frontier. In W. Kellogg and S. Whittaker (eds) *Proceedings of the 2000 ACM conference on Computer Supported Cooperative Work*. Philadelphia, Pennsylvania, United States. ACM. pp. 279-88.

Bertelsen, O. W. and Bødker, S. (2001). Cooperation in massively distributed information spaces. In W. Prinz, M. Jarke, Y. Rogers, K. Schmidt and V. Wulf (eds) *Proceedings of the Seventh European Conference on Computer Supported Cooperative Work, 2001*. Bonn, Germany. Netherlands: Kluwer. pp. 1-17.

Bijker, W. E. (1987). The Social Construction of Bakelite: Toward a Theory of Invention. In W.E. Bijker, T.P. Hughes and T.J. Pinch (eds) *The Social Construction of Technological Systems*. Cambridge, MA: MIT Press. pp. 159-87.

Bijker, W. E. and Law, J. (eds) (1992). *Shaping technology/building society : studies in sociotechnical change*. Cambridge, MA: MIT Press.

Blauner, R. (1964). *Alienation and Freedom: The Factory Worker and His Industry*. Chicago: Chicago University Press.

Blomberg, J., Suchman, L. and Trigg, R. T. (1996). Reflections on a Work-Oriented Design Project. *Human-Computer Interaction* **11**(3). pp. 237-65.

Bloomfield, B. P. and Vurdubakis, T. (1994). Boundary disputes: Negotiating the boundary between the technical and the social in the development of IT systems. *Information Technology & People* **7**(1). pp. 9-24.

Bloor, D. (1992). Left and Right Wittgensteinians. In A. Pickering (ed) *Science as Practice and Culture*. Chicago, IL: The University of Chicago Press. pp. 266-82.

Bloor, D. (1976). *Knowledge and Social Imagery*. London: Routledge & Kegan Paul.

Blythe, M., Wright, P., McCarthy, J., et al. (2006). Theory and method for experience centered design. In *Conference on Human Factors in Computing Systems*. Montréal, Québec, Canada. New York: ACM. pp. 1691-4.

Bødker, S. (1996). *Through the interface: A human activity approach to user interface design*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Bohlin, I. (2004). Communication Regimes in Competition: The Current Transition in Scholarly Communication Seen Trough the Lens of the Sociology of Technology. *Social Studies of Science* **34**(3). pp. 364-91.

Bowers, J., Button, G. and Sharrock, W. (1995). Workflow from within and without. In H. Marmohn, Y. Sundblad and K. Schmidt (eds) *Proceedings of the European Conference on Computer-Supported Cooperative Work, 1995*. Stockholm, Sweden. Springer. pp. 51-66.

Brooks, R. and Stein, L. A. (1994). Building Brains for Bodies. *Autonomous Robots* **1**(1). pp. 7-25.

Büscher, M., Mogensen, P. and Shapiro, D. (2001). Spaces of Practice. In W. Prinz, M. Jarke, Y. Rogers, K. Schmidt and V. Wulf (eds) *Proceedings of the Seventh European Conference on Computer Supported Cooperative Work, 2001*. Bonn, Germany. Kluwer. pp. 139-58.

Button, G. (1997). Cognition in the Wild, Edwin Hutchins (book review). *Computer Supported Cooperative Work* **6**(4). pp. 391-5.

Button, G. (1991a). Introduction: ethnomethodology and the foundational respecification of the human sciences. In G. Button (ed) *Ethnomethodology and the Human Sciences*. Cambridge University Press. pp. 1-8.

Button, G. (ed) (1991b). *Ethnomethodology and the Human Sciences*. Cambridge University Press.

Button, G. and Dourish, P. (1996). Technomethdology: Paradoxes and Possibilities. In *Proceedings of the SIGCHI conference on human factors in computing systems: common ground 1996*. Vancouver, British Columbia, Canada. New York: ACM. pp. 19-26.

Button, G. and Harper, R. (1996). The Relevance of 'Work-Practice' for Design. *Computer Supported Cooperative Work* **4**(4). pp. 263-80.

Button, G. and Sharrock, W. (1998). The Organizational Accountability of Technological Work. *Social Studies of Science* **28**(1). pp. 73-102.

Button, G. and Sharrock, W. (1996). Project Work: The Organisation of Collaborative Design and Development in Software Engineering. *Computer Supported Cooperative Work* **5**(4). pp. 369-86.

Button, G. and Sharrock, W. (1995). The mundane work of reading and writing computer code. In P. ten Havé and G. Psathas (eds) *Situated Order: Studies in the Social Organisation of Talk and Embodied Activities*. Washington DC: University Press of America. pp. 231-58.

Callon, M. (1987). Society in the Making: the Study of Technology as a Tool for Sociological Analysis. In W.E. Bijker, T.P. Hughes and T.J. Pinch (eds) *The Social Construction of Technological Systems*. Cambridge, MA: MIT Press. pp. 83-103.

Carroll, J. M. (ed) (2003). *HCI models, theories, and frameworks: toward a multidisciplinary science*. Elsevier.

Coffey, A. (1999). *The Ethnographic Self: Fieldwork and the Representation of Identity*. London: Sage.

Coffin, J. (2006). Analysis of open source principles in diverse collaborative communities. *First Monday* **11**(6). p. 7. http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/viewArticle/1342/1262

Coleman, J. S. (1968). Studies in Ethnomethodology. by Harold Garfinkel (review). *American Sociological Review* **33**(1). pp. 126-30.

Collins, H. M. (1992). *Changing Order: Replication and Induction in Scientific Practice (with a new Afterword)*. Chicago University Press.

Collins, H. M. (1990). *Artificial experts: social knowledge and intelligent machines*. MIT Press.

Collins, H. M. (1983). An empirical relativist programme in the sociology of scientific knowledge. In K.D. Knorr-Cetina and M.J. Mulkay (eds) *Science Observed: Perspectives on the Social Studies of Science*. Beverly Hills: Sage. pp. 85-113.

Collins, H. M. (1981). Stages in the empirical programme of relativism. *Social Studies of Science* **11**(1). pp. 3-10.

Collins, H. M. and Pinch, T. J. (1998). *The golem: what you should know about science (2nd edition) A Canto Book Series*. Cambridge University Press.

Crabtree, A. (2004). Taking technomethodology seriously: hybrid change in the ethnomethodology-design relationship. *European Journal of Information Systems* **13**(3). p. 195–209.

Crabtree, A. (2001). Doing workplace studies: praxiological accounts - lebenswelt pairs. *TeamEthno Online* **1**(1). p. 17. http://www.mrl.nott.ac.uk/~axc/documents/papers/TEO01.pdf

Crowston, K., Li, Q., Wei, K., et al. (2007). Self-organization of teams for free/libre open source software development. Information and Software Technology. *Information and Software Technology* **49**(6). pp. 564-75.

Denzin, N. K. (1969). Symbolic Interactionism and Ethnomethodology: A Proposed Synthesis. *American Sociological Review* **34**(6). pp. 922-34.

Denzin, N. K. and Lincoln, Y.S. (eds) (2000). *Handbook of Qualitative Research (2nd edition)*. Thousand Oaks, CA: Sage.

Descartes, R. (1991[1637]). *Discours de la méthode (translated into Icelandic by Magnús G. Jónsson w/introduction by Þorsteinn Gylfason)*. Reykjavík: Hið Íslenska Bókmenntafélag.

Dix, A. (1997). Challenges for Cooperative Work on the Web: An Analytical Approach. *Computer Supported Cooperative Work* **6**(2-3). pp. 135-56.

Dourish, P. and Button, G. (1998). On "Technomethodology": Foundational Relationships Between Ethnomethodology and System Design. *Human-Computer Interaction* **13**(4). pp. 395-432.

Downey, G. L. (1998). *The Machine in Me: An Anthropologist Sits Among Computer Engineers*. New York: Routledge.

Dreyfus, H. L. (2007). Why Heideggerian AI Failed and How Fixing it Would Require Making it More Heideggerian. *Philosophical Psychology* **20**(2). pp. 247-68.

Dreyfus, H. L. (1992). *What Computers Still Can't Do: A Critique of Artificial Reason*. MIT Press.

Ducheneaut, N. (2005). Socialization in an Open Source Software Community: A Sociotechnical Analysis. *Computer Supported Cooperative Work* **4**(4). pp. 323-68.

Durkheim, E. (1985[1938]). *The Rules of Sociological Method*. Glencoe IL: Free Press.

Durkheim, E. (1984[1893]). *The Division of Labor in Society (Translated by W.D. Hall)*. Palgrave Macmillan.

Durkheim, E. (1972). *Selected Writings (edited and translated by A. Giddens)*. Cambridge University Press.

Eggen, B. and Kuffin, S. (2006). Experience Design. In E. Aarts and J. Encarnação (eds) *True Visions: The Emergence of Ambient Intelligence*. Springer-Verlag. pp. 361-78.

Ehrlich, K. and Cash, D. (1999). The Invisible World of Intermediaries: A Cautionary Tale. *Computer Supported Cooperative Work* **8**(1-2). pp. 147-67.

Elzen, B. (1986). Two ultracentrifuges: A comparative study of the social construction of artifacts. *Social Studies of Science* 16(4). pp. 621-62.

Engeström, Y. (2000). Activity theory as a framework for analyzing and redesigning work. *Ergonomics* **43**(7). pp. 960-74.

Engeström, Y. (1999). Expansive Visibilization of Work: An Activity-Theoretical Perspective. *Computer Supported Cooperative Work* **8**(1-2). pp. 63-93.

Engeström, Y. (1996). Interobjectivity, ideality, and dialectics. *Mind, Culture and Activity* **3**(4). pp. 259-65.

Fairclough, N. (2003). *Analysing Discourse: Textual Analysis for Social Research*. London: Routledge.

Feyerabend, P. K. (1975). *Against Method: Outline of an Anarchistic Theory of Knowledge*. New York: Humanities Press.

Forsythe, D. E. (2001). *Studying Those Who Study Us: An anthropologist in the world of artificial intelligence (edited w/ introduction by David J. Hess)*. Stanford, CA: Stanford University Press.

Forsythe, D. E. (1999). "It's Just a Matter of Common Sense": Ethnography as Invisible Work. *Computer Supported Cooperative Work* **8**(1-2). pp. 127-45.

Forsythe, D. E. (1995). Using Ethnography in the Design of an Explanation System. *Expert Systems With Applications* **8**(4). pp. 403-17.

Garfinkel, H. (2002). *Ethnomethodology's Program: Working out Durkheim's Aphorism*. Lanham, MD: Rowan & Littlefield.

Garfinkel, H. (1996). Ethnomethodology's program. *Social Psychology Quarterly* **59**(1). pp. 5-21.

Garfinkel, H. (1991). Respecification: evidence for locally produced, naturally accountable phenomena of order, logic, reason, meaning, method, etc. in and as of the essential haecceity of immortal ordinary society (I) - an announcement of studies. In G. Button (ed) *Ethnomethodology and the human sciences*. Cambridge University Press. pp. 10-9.

Garfinkel, H. (1988). Evidence for Locally Produced, Naturally Accountable Phenomena of Order, Logic, Reason, Meaning, Method, etc. In and as of the Essential Quiddity of Immortal Ordinary Society, (I of IV): An Announcement of Studies. *Sociological Theory* **6**(1). pp. 103-9.

Garfinkel, H. (1967). *Studies in Ethnomethodology*. Cambridge, UK: Polity Press.

Garfinkel, H. (ed) (1986). *Ethnomethodological Studies of Work: Studies in Ethnomethodology*. Routledge & Keagan Paul.

Garfinkel, H. and Sacks, H. (1970). On formal structures of practical actions. In J. McKinney and E. Tiryakian (eds) *Theoretical sociology: Perspectives and development*. New York: Appleton Century Crofts. pp. 337-66.

Garfinkel, H., Lynch, M. and Livingston, E. (1981). The work of a discovering science construed with materials for the optically discovered pulsar. *Philosophy of the Social Sciences* **11**(). pp. 131-58.

Garrety, K. and Badham, R. (2004). User-Centered Design and the Normative Politics of Technology. *Science Technology and Human Values* **29**(2). pp. 191-212.

Gerson, E. M. and Star, S. L. (1986). Analyzing Due Process in the Workplace. *ACM Transactions on Office Information Systems* **4**(3). pp. 257-70.

Gerth, H. H. and Mills, C. W. (1948). Introduction: The Man and His Work, III. Intellectual Orientations. In H.H. Gerth and C.W. Mills (eds) *Max Weber, Essays in Sociology (translated w/ introduction)*. London: Routledge & Kegan Paul. pp. 45-74.

Giddens, A. (2000). Globalization and Communication. *Public lecture at the Annenberg's School for Communication, University of Southern California*.

Giddens, A. (1999). *Runaway World: How Globalization is Reshaping Our Lives*. London: Profile Books.

Giddens, A. (1995). *Politics, sociology and social theory: encounters with classical and contemporary social thought*. Stanford University Press.

Giddens, A. (1986). *The Constitution of Society: Outline of the Theory of Structuration*. University of California Press.

Giddens, A. (1972). Introduction. In A. Giddens (ed) *Emile Durkheim: Selected Writings (edited and translated by A. Giddens)*. Cambridge University Press. .

Giddens, A. and Turner, J. (eds) (1987). *Sociological Theory Today*. Oxford: Polity.

Ginsparg, P. (2003). *Can Peer Review be better Focused?*. Published on arXiv.org. http://arxiv.org/blurb/pg02pr.html

Ginsparg, P. (1996). Winners and Losers in the Global Research Village. *Invited talk at UNESCO conference on Scientific Publishing, Paris, 19-23 Feb 1996.* http://arxiv.org/blurb/pg96unesco.html

Glaser, B. G. and Strauss, A. L. (1968). *The Discovery of Grounded Theory*. London: Weidenfeld & Nicolson.

Goffman, E. (1956). *Presentation of Self in Society*. University of Chicago Press.

Goldthorpe, J. H. (1966). Attitudes and Behaviour of Car-Assembly Workers: A Deviant Case and a Theoretical Critique. *British Journal of Sociology* **17**(3). pp. 227-44.

Goodwin, C. (1994). Professional Vision. *American Anthropologist* **96**(3). pp. 606-33.

Görling, S. (2007). Open Source Athletes. *First Monday* **12**(4). p. 8. http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/viewArticle/1770/1650

Grasso, A., Meunier, J., Paganini, D., et al. (1997). Distributed Coordination and Workflow on the World Wide Web. *Computer Supported Cooperative Work* **6**(2-3). pp. 175-200.

Gross, T. and Prinz, W. (2003). Awareness in Context: A Light-Weight Approach. In K. Kuutti, E.H. Karsten, G. Fitzpatrick, P. Dourish and K. Schmidt (eds) *Proceedings of the Eighth European Conference on Computer Supported Cooperative Work, 2001*. Bonn, Germany. Kluwer. pp. 295-314.

Guédon, J. (2001). *In Oldenburg's Long Shadow: Librarians, Research Scientists, Publishers, and the Control of Scientific Publishing*. Association of Research Libraries (ARL). Available online at http://www.arl.org/resources/pubs/mmproceedings/138guedon.shtml.

Gunnarsdóttir, K. (2005). Scientific Journal Publications: On the Role of Electronic Preprint Exchange in the Distribution of Scientific Literature. *Social Studies of Science* **35**(4). pp. 549-79.

Gutwin, C., Penner, R. and Schneider, K. (2004). Group Awareness in Distributed Software Development. In J. Herbsleb and G. Olson (eds) *Proceedings of the 2004 ACM conference on Computer Supported Cooperative Work*. Chicago, Illinois, USA. New York: ACM. pp. 72-81.

Gyert, R. M. and March, J. G. (1963). *A Behavioural Theory of the Firm*. Englewood Cliffs, NJ: Prentice-Hall.

Hacking, I. (2002). *Historical Ontology*. Harvard University Press.

Halverson, C. A. (2002). Activity Theory and Distributed Cognition: Or What Does CSCW Need to DO with Theories?. *Computer Supported Cooperative Work* **11**(1-2). pp. 243-67.

Hammersley, M. and Atkinson, P. (1994). *Ethnography, principles in practice*. London: Routledge.

Harnad, S. (2001). *For Whom the Gate Tolls? How and Why to Free the Refereed Research Literature Online Through Author/Institution Self-Archiving, Now*. University of Southampton, Intelligence/Agents/Multimedia Group. http://users.ecs.soton.ac.uk/harnad/Tp/resolution.htm

Harnad, S. (1991). Post-Gutenberg Galaxy: The Fourth Revolution in the Means of Production of Knowledge. *Public-Access Computer Systems Review* **2**(1). pp. 39-53. http://cogprints.org/1580/00/harnad91.postgutenberg.html

Heath, C. and Luff, P. (2000). *Technology in Action*. Cambridge University Press.

Heath, C. and Luff, P. (1998). Convergent Activities: Line Control and passenger Information on London Underground. In Y. Engeström and D. Middleton (eds) *Cognition and Communication at Work*. Cambridge University Press. pp. 96-129.

Heath, C., Knoblauch, H. and Luff, P. (2000). Technology and social interaction: the emergence of 'workplace studies'. *British Journal of Sociology* **51**(2). pp. 299-320.

Heath, C., Luff, P., Kuzuoka, H., et al. (2001). Creating Coherent Environments for Collaboration. In W. Prinz, M. Jarke, Y. Rogers, K. Schmidt and V. Wulf (eds) *Proceedings of the Seventh European Conference on Computer Supported Cooperative Work, 2001*. Bonn, Germany. Netherlands: Kluwer. pp. 119-38.

Heath, C., Svensson, S. M., Hindmarsh, J., et al. (2002). Configuring Awareness. *Computer Supported Cooperative Work* **11**(3). p. 317–347.

Hemmings, T., Crabtree, A., Rodden, T., et al. (2002). Probing the Probes. In T. Binder, J. Gregory and I. Wagner (eds) *Proceedings of the Participatory Design Conference, 23-25 June*. Malmö, Sweden. PDC 2002. pp. 42-50. http://www.mrl.nott.ac.uk/~axc/documents/papers/PDC02.pdf

Herring, S. (1999). Interactional Coherence in CMC. *Journal of Computer Mediated Communication* **4**(4). p. 24.

Hess, D. (2001). Ethnography and the Development of Science and Technology Studies. In P. Atkinson, A. Coffey, S. Delamont, J. Lofland and L. Lofland (eds) *Handbook of Ethnography*. London: Sage. .

Hindmarsh, J., Fraser, M., Heath, C., et al. (1998). Fragmented Interaction: Establishing mutual orientation in virtual environments. In *Proceedings of the 1998 ACM conference on Computer Supported Cooperative Work*. Seattle, Washington, United States. New York: ACM. pp. 217-26.

Hine, C. (2000). *Virtual Ethnography*. London, Thousand Oaks: Sage.

Hollan, J., Hutchins, E. and Kirsh, D. (2000). Distributed cognition: toward a new foundation for human-computer interaction research. *ACM Transactions on Computer-Human Interaction (TOCHI)* **7**(2). pp. 174-96.

Hughes, J. A., Randall, D. and Shapiro, D. (1993). From ethnographic record to system design. *International Journal of CSCW* **1**(3). pp. 123-41.

Hughes, T. P. (1986). The Seamless Web: Technology, Science Etcetera. *Social Studies of Science* **16**(2). pp. 281-92.

Hughes, T. P. (1985). Edison the electric light. In D. MacKenzie and J. Wajcman (eds) *The Social Shaping of Technology: How the refrigerator got its hum*. Philadelphia: Open University Press. pp. 39-52.

Hutchins, E. (1995b). *Cognition in the wild*. Cambridge, Mass.: MIT Press.

Hutchins, E. (1995a). How a cockpit remembers its speeds. *Cognitive Science* **19**(3). pp. 265-88.

Hvannberg, E. Þ. (2006). Experience Research. In E. Aarts and J. Encarnação (eds) *True Visions: The Emergence of Ambient Intelligence*. Springer-Verlag. pp. 379-94.

Iannacci, F. and Mitleton-Kelly, E. (2005). Beyond markets and firms: The emergence of Open Source networks. *First Monday* **10**(5). p. 12. http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/viewArticle/1237/1157

Jones, S. G. (ed) (1999). *Doing Internet Research: Critical Issues and Methods for Examining the Net*. London: Sage.

Jones, S. K. (ed) (1981). *Information Retrieval Experiments*. London: Butterworks.

Jordan, K. and Lynch, M. (1992). The Sociology of a Genetic Engineering Technique: Ritual and Rationality in the Performance of the “Plasmid Prep”. In A.E. Clark and J.H. Fujimura (eds) *The Right Tools For The Job*. Princeton NJ: Princeton University Press. pp. 77-114.

Kaptelinin, V. and Nardi, B. A. (2006). *Acting with Technology: Activity Theory and Interaction Design*. MIT Press.

Kling, R. (1991). Computerization and Social Transformations. *Science Technology & Human Values* **16**(3). pp. 342-67.

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

Knorr-Cetina, K. D. (1981). The Micro-sociological Challenge of Macro-sociology: Towards a Reconstruction of Social Theory and Methodology. In K.D. Knorr-Cetina and A.V. Cicourel (eds) *Advances in Social Theory and Methodology: Towards an Integration of Micro- and Macro-sociologies*. London: Routledge and Kegan Paul. pp. 1-47.

Kuhn, T. S. (1962). *The Structure of Scientific Revolutions*. The University of Chicago Press.

Kuperberg, G. (2002). *Scholarly Mathematical Communication at a Crossroads*. Published on arXiv.org. http://arxiv.org/math.HO/0210144

Kuutti, K. (1996). Activity theory as a potential framework for human-computer interaction research. In B.A. Nardi (ed) *Context and Consciousness*. Cambridge MA: MIT Press. pp. 17-44.

Latour, B. (2005). *Reassembling the Social: An Introduction to Actor-Network-Theory*. Oxford University Press.

Latour, B. (1993). *We Have Never Been Modern (translated by Catherine Porter)*. Harvard University Press.

Latour, B. (1991). Technology is Society Made Durable. In J. Law (ed) *Sociology of Monsters: Essays on power, technology and domination*. London and Yew York: Routledge. pp. 103-31.

Latour, B. (1988). Mixing Humans and Nonhumans Together: The Sociology of a Door-Closer. *Social Problems* **35**(3). pp. 298-310.

Latour, B. (1987). *Science in Action: How to follow scientists and engineers through society*. Cambridge MA: Harvard University Press.

Law, J. (ed) (1991). *A Sociology of Monsters? Essays on Power, Technology and Domination (Sociological Review Monograph)*. London, Routledge.

Law, J. and Mol, A. (2002a). Complexities: An Introduction. In J. Law and A. Mol (eds) *Complexities: Social Studies of Knowledge Practices*. Duke University Press. pp. 1-22.

Law, J. and Mol, A. (eds) (2002b). *Complexities: Social Studies of Knowledge Practices*. Duke University Press.

Lazarsfeld, P. F. (1958). Evidence and Inference in Social Research. *Daedalus (On Evidence and Inference)* **87**(4). pp. 99-130.

Lazarsfeld, P. F. and Rosenberg, M. (eds) (1955). *The Language of Social Research*. New York: The Free Press.

Lehmann, F. (2004). FLOSS developers as a social formation. *First Monday* **9**(11). p. 11. http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/viewArticle/1186/1106

Lévi-Strauss, C. (1963). *Structural Anthropology*. New York: Basil Books.

Luff, P., Hindmarsh, J. and Heath, C. (eds) (2000). *Workplace Studies: Recovering Work Practice and Informing System Design*. Cambridge University Press.

Lynch, M. (1999). Silence in Context: Ethnomethodology and Social Theory. *Human Studies* **22**(2-4). pp. 211-33.

Lynch, M. (1993). *Scientific practice and ordinary action: Ethnomethodology and social studies of science*. Cambridge University Press.

Lynch, M. (1992b). Extending Wittgenstein: The Pivotal move from Epistemology to the Sociology of Science. In A. Pickering (ed) *Science as Practice and Culture*. Chicago, IL: The University of Chicago Press. pp. 215-65.

Lynch, M. (1992a). From the "Will to Theory" to the Discursive Collage: A Reply to Bloor's "Left and Right Wittgensteinians. In A. Pickering (ed) *Science as Practice and Culture*. Chicago, IL: The University of Chicago Press. pp. 383-300.

Lynch, M. (1985). *Art and Artifact in Laboratory Science: A Study of Shop Work and Shop Talk in a Research Laboratory*. Routledge Kegan & Paul.

Lynch, M. (1982). Technical Work and Critical Inquiry: Investigations in a Scientific Laboratory. *Social Studies of Science* **12**(4). pp. 499-533.

Lynch, M. and Bogen, D. (1996). *The Spectacle of History: Speech, Text and Memory at the Iran-Contra Hearings (Post-Contemporary Interventions)*. Duke University Press.

Mackay, H., Carne, C., Beynon-Davies, P., et al. (2000). Reconfiguring the User: Using Rapid Application Development. *Social Studies of Science* **30**(5). pp. 737-57.

MacKenzie, D. (1990). *Inventing accuracy: A historical sociology of nuclear missile guidance*. Cambridge, MA: MIT Press.

Markham, A. N. (1998). *Life Online: Researching Real Experience in Virtual Space*. London: Altamira Press, Sage.

Marx, K. and Engels, F. (1998b). *The German Ideology: Including Theses on Feuerbach and an Introduction to the Critique of Political Economy*. Prometheus Books.

Marx, K. and Engels, F. (1998a). *The Communist Manifesto (edited w/ introduction by Martin Malia)*. New York: Penguin Group.

McCarthy, J. and Wright, P. (2004). *Technology as Experience*. The MIT Press.

Merton, R. K. (1959). Introduction: Notes on Problem-Finding in Sociology. In R.K. Merton, L. Broom and L.S. Cottrell Jr (eds) *Sociology Today: Problems and Prospects*. New York: Basic Books. pp. ix-xxxiv.

Merton, R. K. (1957). *Social Theory and Social Structure (revised and enlarged)*. Glencoe IL: The Free Press.

Merton, R. K., Broom, L. and Cottrell Jr, L.S. (eds) (1959). *Sociology Today: Problems and Prospects*. New York: Basic Books.

Miles, W. D. (1992). *A History of the National Library of Medicine: The Nation's Treasury of Medical Knowledge*. U.S Government printing Office.

Misa, T. J. (1992). Controversy and closure in technological change: Constructing steel. In W. Bijker and J. Law (eds) *Shaping technology/building society : studies in sociotechnical change*. Cambridge, MA: MIT Press. pp. 109-139

Mooers, C. N. (1950). Information Retrieval Viewed as Temporal Signalling. In *Proceedings of the International Congress of Mathematicians, 30 August - 06 September 1950*. Harvard University. Cambridge, Massachusetts (Vol.1). pp. 572-3.

Morgan, G. (1986). *Images of Organization*. Beverly Hills: Sage.

Muller, M. J. and Gruen, D. M. (2005). Working Together Inside an Emailbox. In H.W. Gellersen, K. Schmidt, M. Beaudouin-Lafon and W. Mackay (eds) *Proceedings of the Ninth European Conference on Computer-Supported Cooperative Work, 2005*. Paris, France. Springer. pp. 103-22.

Nardi, B. A. (2002). Coda and Response to Christine Halverson. *Computer Supported Cooperative Work* **11**(1-2). p. 269–275.

Nardi, B. A. (1996a). Activity theory and human-computer interaction. In B.A. Nardi (ed) *Context and Consciousness: Activity Theory and Human-Computer Interaction*. MIT Press. pp. 7-10.

Nardi, B. A. (ed) (1996b). *Context and Consciousness: Activity Theory and Human-Computer Interaction*. MIT Press.

Nardi, B. A., Whittaker, S. and Bradner, E. (2000). Interaction and Outeraction: Instant Messaging in Action. In W. Kellogg and S. Whittaker (eds) *Proceedings of the 2000 ACM conference on Computer Supported Cooperative Work*. Philadelphia, PA. New York: ACM. pp. 79-88.

Ngwenyama, O. K. and Lyytinen, K. J. (1997). Groupware Envirnonments as Action Constitutive Resources: A Social Action Framework for Analyzing Groupware Technologies. *Computer Supported Cooperative Work* **6**(1). pp. 71-93.

Nichols, D. M. and Twidale, M. B. (1999). Computer Supported Cooperative Work and Libraries. *Vine: Special Issue on Virtual Communities and Information Services* **109**(5). pp. 10-5.

Nuvolari, A. (2005). Open source software development: Some historical perspectives. *First Monday* **10**(10). p. 15. http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/viewArticle/1284/1204

O'Connell, H. B. (2002). Physicists Thriving with Paperless Publishing. *High Energy Physics Webzine*, *6*. http://library.cern.ch/HEPLW/6/papers/3/

O'Connell, J. (1993). Metrology: The Creation of Universality by the Circulation of Particulars. *Social Studies of Science* **23**(1). pp. 129-73.

Odlyzko, A. M. (2003). Peer and Non-Peer Review. In F. Godlee and T. Jefferson (eds) *Peer Review in Health Sciences (2nd edition)*. BMJ Books. pp. 309-11. http://www.dtc.umn.edu/~odlyzko/doc/peer.review.txt

Odlyzko, A. M. (2000). The future of scientific communication. In P. Schroeder and P. Wouters (eds) *Access to Publicly Financed Research: The Global Research Village III*. Amsterdam: NIWI. pp. 273-8. http://www.dtc.umn.edu/~odlyzko/doc/future.scientific.comm.pdf

Okerson, A. S. and O'Donnell, J. J. (1995). *A Subversive Proposal for Electronic Publishing: An Internet Discussion about Scientific and Scholarly Journals and Their Future*. Office of Scientific & Academic Publishing, Association of Research Libraries (ARL). http://www.arl.org/sc/subversive/

Orlikowski, W. J. (1992). Learning from Notes: Organizational Issues in Groupware Implementation. In M. Mantel and R. Baecker (eds) *Proceedings of the 1992 ACM conference on Computer-supported cooperative work*. Toronto, Ontario, Canada. New York: ACM. pp. 362-9.

Østerlie, T. (2004). In the Network: Distributed Control in Gentoo Linux. In J. Feller, B. Fitzgerald, S. Hissam and K. Lakhani (eds) *Proceedings of the 4th International Workshop on Open Source Software Engineering: Collaboration, Conflict and Control*. Edinburgh, Scotland. opensource.ucc.ie. pp. 76-81. http://opensource.ucc.ie/icse2004/Workshop_on_OSS_Engineering_2004.pdf

Oudshoorn, N. and Pinch, T. (eds) (2003). *How Users Matter: The Co-Construction of Users and Technology*. Massachusetts: MIT Press.

Oudshoorn, N., Rommes, E. and Stienstra, M. (2004). Configuring the User as Everybody: Gender and Design Cultures in Information and Communication Technologies. *Science Technology Human Values* **29**(1). pp. 30-63.

Parsons, T. (1937). *The Structure of Social Action*. New York: McGraw-Hill.

Pickering, A. (ed) (1992). *Science as Practice and Culture*. Chicago: University of Chicago Press.

Pinch, T. J. and Bijker, W. E. (1984). The social construction of facts and artefacs: Or how the sociology of science and the sociology of technology might benefit each other. *Social Studies of Science* **14**(3). pp. 399-441.

Pinfield, S. (2001). How do Physicists Use an E-Print Archive? Implications for Institutional E-Print Services. *D-Lib Magazine* **7**(12). p. 10.

Polanyi, M. (1962). *Personal Knowledge: Towards a Post-Critical Philosophy (corrected edition)*. Chicago IL: The University of Chicago Press.

Pollock, N. (2005). When Is a Work-Around? Conflict and Negotiation in Computer Systems Development. *Science, Technology and Human Values* **30**(4). pp. 496-514.

Prinz, W. (1999). NESSIE: An awareness environment for cooperative settings. In S. Bødker, M. Kyng and K. Schmidt (eds) *Proceedings of the Sixth European Conference on Computer-Supported Cooperative Work, 1999*. Copenhagen, Denmark. Netherlands: Kluwer. pp. 391-410.

Ramduny, D., Dix, A. and Rodden, T. (1998). Exploring the design space for notification servers. In S. Poltrock and J. Grudin (eds) *Proceedings of the 1998 ACM conference on Computer Supported Cooperative Work*. Seattle, Washington, United States. New York: ACM. pp. 227-35.

Raymond, E. S. (1998b). The Cathedral and the Bazaar. *First Monday* **3**(3). p. 16. http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/viewArticle/578/499

Raymond, E. S. (1998a). Homesteading the Noosphere. *First Monday* **3**(10). p. 19. http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/viewArticle/621/542

Reed, M. and Hughes, M. (eds) (1992). *Rethinking organization: new directions in organization theory and analysis*. Sage Publications.

Reeves, K. (1967). Constrained and Facilitated Behaviour. *British Journal of Industrial Relations* **5**(2). pp. 145-61.

Rheingold, H. (2000). *The Virtual Community: Homesteading the on the Electronic Frontier*. MIT Press.

Rogers, Y. and Ellis, J. (1994). Distributed Cognition: an alternative framework for analysing and explaining collaborative working. *Journal of Information Technology* **9**(2). pp. 119-28.

Rogers, Y., Scaife, M., Harris, E., et al. (2002). Things aren't what they seem to be: innovation through technology inspiration. In *Proceedings of the Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques*. London, England. New York: ACM. pp. 373-8.

Ruhleder, K. (1995). Reconstructing Artifacts, Reconstructing Work: From Textual Edition to On-Line Databank. *Science, Technology & Human Values* **20**(1). pp. 39-64.

Ruhleder, K. and Jordan, B. (1999). Meaning-Making Across Remote Sites: How Delays in Transmission Affect Interaction. In S. Bødker, M. Kyng and K. Schmidt (eds) *Proceedings of the Sixth European Conference on Computer-Supported Cooperative Work, 1999*. Copenhagen, Denmark. Netherlands: Kluwer. pp. 411-29.

Ryle, G. (1954). *Dilemmas (Tarner lectures, Vol.115 of Philosophy)*. Cambridge University Press.

Ryle, G. (1949). *Concept of Mind*. London: Hutchinson.

Sack, W., Détienne, F., Ducheneaut, N., et al. (2006). A methodological Framework for Socio-Cognitive Analysis of Collaborative Design of Open Source Software. *Computer Supported Cooperative Work* **15**(2-3). pp. 229-50.

Sayles, L. R. (1958). *Behavior of Industrial Work Group*. New York: Wiley.

Schaffer, S. (1998). The Leviathan of Parsonstown: Literary Technology and Scientific Representation. In T. Lenoir (ed) *Instituting Science: The Cultural Production of Scientific Disciplines*. Stanford University Press. pp. 182-222.

Schatzki, T. R., Knorr-Cetina, K. and von Savigny, E. (eds) (2001). *The Practice Turn in Contemporary Theory*. London: Routledge.

Schmidt, K. and Bannon, L. (1992). Taking CSCW seriously: Supporting articulation work. *Computer Supported Cooperative Work* **1**(1). pp. 7-40.

Schmidt, K. and Simone, C. (1996). Coordination mechanisms: towards a conceptual foundation of CSCW systems design. *Computer Supported Cooperative Work* **5**(2-3). pp. 155-200.

Schütz, A. (1943). The Problem of Rationality in the Social World. *Economica* **10**(38). pp. 130-49.

Schwartz Cowan, R. (1987). The Consumption Junction: A Proposal for Research Strategies in the Sociology of Technology. In W.E. Bijker, T.P. Hughes and T.J. Pinch (eds) *The Social Construction of Technological Systems*. Cambridge, MA: MIT Press. pp. 261-80.

Scott, J. (1990). *A Matter of Record. Documentary Sources in Social Research*. Cambridge, UK: Polity Press.

Sengers, P. and Gaver, B. (2006). Staying open to interpretation: engaging multiple meanings in design and evaluation. In *Proceedings of the 6th ACM conference on Designing Interactive systems, 26-28 June 2006*. University Park, PA. New York: ACM. pp. 99-108.

Shane, P. M. (ed) (2004). *Democracy online: the prospects for political renewal through the Internet*. New York: Routledge.

Sharrock, W. and Anderson, B. (1991). Epistemology: professional scepticism. In G. Button (ed) *Ethnomethodology and the human sciences*. Cambridge University Press. pp. 51-76.

Sharrock, W. and Anderson, R. J. (1987). The Definition of Alternatives: Some Sources of Confusion in Interdisciplinary Discussion. In G. Button and J.R.E. Lee (eds) *Talk and Social Organisation*. Clevedon: Multilingual Matters. pp. 290-321.

Sharrock, W. and Button, G. (1999). Do the right thing! Rule finitism, rule scepticism and rule following. *Human Studies* **22**(2-4). pp. 193-210.

Sharrock, W. and Button, G. (1991). The social actor: social action in real time. In G. Button (ed) *Ethnomethodology and the Human Sciences*. Cambridge University Press. pp. 137-75.

Silverman, D. (2006). *Interpreting Qualitative Data: Methods for Analyzing Talk, Text and Interaction*. Sage Publications.

Silverman, D. (1997). Studying Organizational Interaction: Ethnomethodology's Contribution to the New Institutionalism. *Administrative Theory and Praxis* **19**(2). pp. 178-95.

Silverman, D. (1994). On throwing away ladders: re-writing the theory of organizations. In J. Hassard and M. Parker (eds) *Towards a new theory of organizations*. London: Routledge. pp. 1-23.

Silverman, D. (1970). *The Theory of Organisations: A Sociological Framework*. London: Heinemann.

Smith, M., Cadiz, J. J. and Burkhalter, B. (2000). Conversation Trees and Threaded Chats. In W. Kellogg and S. Whittaker (eds) *Proceedings of the 2000 ACM conference on Computer Supported Cooperative Work*. Philadelphia, PA. New York: ACM. pp. 97-105.

Solove, D. J. (2004). *The Digital Person: Technology and Privacy in the Information Age*. New York University Press.

Stake, R. E. (2000). Case Studies. In N.K. Denzin and Y.S. Lincoln (eds) *Handbook of Qualitative Research (2nd edition)*. Thousand Oaks, CA: Sage. pp. 435-54.

Star, S. L. (1991). Invisible Work and Silenced Dialogues in Knowledge Representation. In I.V. Eriksson, B.A. Kitchenham and K.G. Tijdens (eds) *Women, Work and Computerization: Understanding and Overcoming Bias in Work and Education*. Amsterdam: Elsevier Science Publishers. pp. 81-92.

Star, S. L. 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**(3). pp. 387-420.

Suchman, L. A. (2007). *Human-Machine Reconfigurations: Plans and Situated Actions (2nd edition)*. Cambridge University Press.

Suchman, L. A. (2000b). Embodied Practices of Engineering Work. *Mind, Culture and Activity* **7**(1&2). pp. 4-18.

Suchman, L. A. (2000a). Organizing Alignment: A Case of Bridge-building. *Organization* **7**(2). pp. 311-27.

Suchman, L. A. (1998). Do Categories Have Politics? The Language/Acton Perspective Reconsidered. In B. Friedman (ed) *Human Values and the Design of Computer Technology*. Stanford, CA: CSLI Publications. pp. 91-105.

Suchman, L. A. (1997). Centers of Coordination: A Case and Some Themes. In L.B. Resnick, R. Säljö, C. Pontecorvo and B. Burge (eds) *Discourse, Tools, and Reasoning: Essays on Situated Cognition (NATO ASI Series F. Computer and Systems Sciences, Vol. 160)*. Springer. pp. 41-62.

Suchman, L. A. (1995). Making work visible. *Communications of the ACM* **38**(9). pp. 56-64.

Suchman, L. A. (1987). *Plans and Situated Action: The Problem of Human-Computer Communication (Learning in Doing: Social, Cognitive & Computational Perspectives)*. Cambridge University Press.

Suchman, L. A. (1983). Office Procedure as Practical Action: Models of Work and System Design. *ACM Transactions on Office Information Systems* **1**(4): 320-328.

Swanson, D. R. (1988). Historical Note: Information Retrieval and the Future of an Illusion. *JASIS* **39**(2). pp. 92-9.

Ten Have, P. (2004). *Understanding Qualitative Research and Ethnomethodology*. Sage Publications.

Traweek, S. (1988). *Beamtimes and lifetimes: the world of high energy physicists*. Harvard University Press.

Trigg, R. H., Blomberg, J. and Suchman, L. A. (1999). Moving document collections online: The evolution of a shared repository. In S. Bødker, M. Kyng and K. Schmidt (eds) *Proceedings of the Sixth European Conference on Computer-Supported Cooperative Work, 1999*. Copenhagen, Denmark. Kluwer. pp. 331-50.

Tuomi, I. (2001). Internet, Innovation, and Open Source: Actors in the Network. *First Monday* **6**(1). p. 17. http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/viewArticle/824/733

Twidale, M. B. and Marty, P. F. (2000). Coping with Errors: the Importance of Process Data in Robust Sociotechnical Systems. In W. Kellogg and S. Whittaker (eds) *Proceedings of the 2000 ACM conference on Computer Supported Cooperative Work, 2000*. Philadelphia, Pennsylvania, United States. New York: ACM. pp. 269-78.

Vygotsky, L. (1978). *Mind in Society: Development of Higher Psychological Processes (new edition)*. Harvard University Press.

Walker, C. R. and Guest, R. H. (1952). *The Man on the Assembly Line*. Cambridge, MA: Harvard University Press.

Warner, S. (2001). Open Archives Initiative protocol development and implementation at arXiv. *Expanded version of talk presented at Open Archives Initiative Open Meeting in Washington DC, USA on 23 January 2001*. http://arxiv.org/pdf/cs/0101027v1 Last accessed Mar. 2008

Watson, T. (2002). Professions and Professionalism: Should We Jump Off the Bandwagon, Better to Study Where It Is Going?. *International Journal of Management & Organisation* **32**(2). pp. 93-105.

Weber, M. (1948). *From Max Weber: Essays in Sociology (translated and edited w/ introduction by H. H. Gerth and C. W. Mills)*. London: Routledge & Kegan Paul.

Weber, M. (1947). *The Theory of Social and Economic Organisation (translated by A. M. Henderson and T. Parsons w/ introduction by T. Parsons)*. Glencoe, IL: The Free Press & The Falcon's Wing Press.

Weeler, M. (2005). *Reconstructing the cognitive world: the next step*. MIT Press.

Werdel, J. A. and Adams, S. (1976). U.S. Participation in world information activities. *Bulletin of the American Society for Information Science* **2**(8). pp. 44-6.

Williams, M. L. (2006). *Virtually Criminal: Crime, Deviance and Regulation Online*. London: Routledge.

Winch, P. (1990). *The Idea of a Social Science and its Relation to Philosophy (2nd edition)*. Routledge.

Winner, L. (1993). Upon Opening the Black Box and Finding It Empty: Social Constructivism and the Philosophy of Technology. *Science, Technology, & Human Values* **18**(3). pp. 362-78.

Winograd, T. (1995). Heidegger and the Design of Computer Systems. In A. Feenberg and A. Hannay (eds) *Technology and the Politics of Knowledge*. Bloomington: Indiana University Press. pp. 108-27.

Winograd, T. and Flores, F. (1986). *Understanding computers and cognition: A new foundation for design*. NJ: Ablex, Norwood.

Wittel, A. (2000). Ethnography on the move: From Field to Net to Internet. *Forum: Qualitative Social Research* **1**(1). p. 9. http://www.qualitative-research.net/fqs-texte/1-00/1-00wittel-e.htm

Wittgenstein, L. (1953). *Philosophical Investigations (translated by G. E. M. Anscombe)*. Oxford, UK: Blackwell Publishers.

Woolgar, S. (1991b). The Turn to Technology in Social Studies of Science. *Science, Technology and Human Values* **16**(1). pp. 20-50.

Woolgar, S. (1991a). Configuring the user: the case of usability trials. In J. Law (ed) *A Sociology of Monsters? Essays on Power, Technology and Domination (Sociological Review Monograph)*. London, Routledge. pp. 57-99.

Woolgar, S. (ed) (2002). *Virtual Society?: Technology, Cyberbole, Reality*. Oxford University Press.

Wright, D., Gutwirth, S., Friedewald, M., et al. (2008). *Safeguards in a World of Ambient Intelligence. The International Library of Ethics, Law and Technology, vol.1*. Springer Verlag.

Yamauchi, Y., Yokozawa, M., Shinohara, T., et al. (2000). Collaboration with Lean Media: how open-source software succeeds. In W. Kellogg and S. Whittaker (eds) *Proceedings of the 2000 ACM conference on Computer Supported Cooperative Work*. Philadelphia, PA, United States. New York: ACM. pp. 329-38.

Appendices

APPENDIX I

Glossary of terms

Adapted and localised from wikipedia

backward compatibility

Irrespective of platform (hardware/operating system), a new product is said to be backward compatible when it is able to take the place of an older product, by interoperating with devices that were designed for the older product.

(http://en.wikipedia.org/wiki/Backward_compatibility)

batch job - batch processing

Batch jobs are set up so they can be run to completion without human interaction, so all input data is pre-selected through scripts or command line parameters. This is in contrast to "online" or interactive programs which prompt the user for such input.

(http://en.wikipedia.org/wiki/Batch_job)

compiler

A compiler is a computer program (or set of programs) that translates text written in a computer language (*source language*) into another computer language (*target language*). The original sequence is usually called the *source code* and the output is called *object code*.

(http://en.wikipedia.org/wiki/Compiler)

copyleft - copyleft license

The term is a play on the word copyright and describes the practice of using copyright law to remove restrictions on distributing copies and modified versions of work. An author may, through a copyleft licensing scheme, give every person who receives a copy of the work permission to reproduce, adapt or distribute that work as long as any resulting copies or adaptations are also bound by that same licensing scheme.

(http://en.wikipedia.org/wiki/Copyleft)

daemon

In Unix and other computer multitasking operating systems, a daemon is a computer program that runs in the background, rather than under the direct control of a user; they are usually initiated as processes.

(http://en.wikipedia.org/wiki/Daemon_%28computer_software%29)

data transmission

Sending a stream of bits or bytes from one location to another. Practical examples include moving data from one data storage device to another such as accessing a website, which involves data transfer from web servers to a user's browser. `FTP`, `HTTP`, `BIT-NET` are communications protocols used to transfer data/information/documents from one computer to another through a network, such as over the Internet.

(http://en.wikipedia.org/wiki/Data_transmission - http://en.wikipedia.org/wiki/Communications_protocol#Protocol_families)

database

A computer database is a structured collection of records or data that is stored in a computer system and relies upon software to organize the storage of the data. The software can model the database structure in what are known as database models (or data models).

flat: A flat file database is a file that contains records, and in which each record is specified in a single line. Fields from each record may simply have a fixed width with padding, or may be delimited by whitespace, tabs, commas or other characters. There are no structural relationships. The data are "flat" as in a sheet of paper, in contrast to more complex models such as a relational database.

relational: The basic data structure of the relational model can be thought of as a table, where information about a particular entity (say, an employee) is represented in columns and rows (also called tuples). Thus, the "relation" in a "relational database" refers to the various tables in the database; a relation is a set of tuples. The relational model specifies that the tuples of a relation should have no specific order and that the tuples, in turn, should impose no order on the attributes, however, all relations (and, thus, tables) have to adhere to some basic rules to qualify as relations.

(http://en.wikipedia.org/wiki/Database - http://en.wikipedia.org/wiki/Relational_database - http://en.wikipedia.org/wiki/Flat_file_database)

groupware – collaborative software

Collaborative software is software designed to help people involved in a common task achieve their goals. Collaborative software is the basis for computer supported cooperative work. Such software systems as email, calendaring, text chat, and wiki belong in this category

(http://en.wikipedia.org/wiki/Groupware)

hacker - tinkerer

In one of several meanings of the word in computing, a hacker is a member of the programmer subculture which originated in the 1960s in the United States academia, in particular around the Massachusetts Institute of Technology (MIT). Nowadays, this subculture is mainly known for the free software movement. Hackers follow a spirit of creative playfulness and anti-authoritarianism. They have been described as "a person who enjoys exploring the details of programmable systems and how to stretch their capabilities, as opposed to most users, who prefer to learn only the minimum necessary" or "a person who delights in having an intimate understanding of the internal workings of a system, computers and computer networks in particular."

(http://en.wikipedia.org/wiki/Hacker_%28free_and_open_source_software%29)

header

In information technology, header refers to supplemental data placed at the beginning of a block of data being stored or transmitted, which contain information for the handling of the data block. In graphics file formats, the header might give information about an image's size, resolution, number of colors, and the like.

(http://en.wikipedia.org/wiki/Header_%28information_technology%29)

image - figure formats

Image files are made up of either pixel or vector (geometric) data, which is rasterized to pixels in the display process, with a few exceptions in vector graphic display. The pixels that make up an image are in the form of a grid of columns and rows. Each pixel in an image consists of numbers representing brightness and color.

bitmaps - raster graphics: A raster graphics image or bitmap is a data structure representing a generally rectangular grid of pixels, or points of color, viewable via a monitor, paper, or other display medium. Raster images are stored in image files with varying formats (also known as pixmaps): `GIF`, `JPEG (JPG)`, `PGN`, `PNM`, `TIFF`.

lossy: Lossy compression algorithms (usually for the `JPEG` format) take advantage of the inherent limitations of the human eye and discard information that cannot be seen. Most lossy compression algorithms allow for variable levels of quality (compression) and as these levels are increased, file size is reduced and visual quality lost.

(http://en.wikipedia.org/wiki/Raster_graphics - http://en.wikipedia.org/wiki/Image_file)

lurking - lurker

In internet culture, a lurker is a person who reads discussions on a message board, in a newsgroup, a chatroom, file sharing or other interactive systems, but rarely participates. Lurking is also way to get a feel for what is going on or to quietly monitor without stepping in unless input is needed.

(http://en.wikipedia.org/wiki/Lurking)

metadata

Metadata are data about data. An item of metadata may describe an individual datum, a content item, or a collection of data contents. Metadata required for effective data management varies with the type of data and context of use. In a library, metadata would typically include a description of the content, the author, the publication date and the physical location. In the context of an information system, where the data is the content of the computer files, metadata about an individual data item would typically include the name of the field and its length.

(http://en.wikipedia.org/wiki/Metadata)

parse - parsing

In computer science and linguistics, parsing (more formally: syntactic analysis) is the process of analyzing a sequence of tokens to determine grammatical structure with respect to a given (more or less) formal grammar. A parser is thus one of the components in an interpreter or compiler, where it captures the implied hierarchy of the input text and transforms it into a form suitable for further processing (often some kind of parse tree, abstract syntax tree or other hierarchical structure) and normally checks for syntax errors at the same time.

(http://en.wikipedia.org/wiki/Parsing)

platform

A platform describes some sort of hardware architecture or software framework (including application frameworks), that allows software to run. Typical platforms include an operating system, programming languages and related runtime libraries or graphical user interfaces.

(http://en.wikipedia.org/wiki/Platform_%28computing%29)

postscript

postscript (`PS`) is a page description language and programming language used primarily in the electronic and desktop publishing areas.

encapsulated postscript (`EPS`): A postscript document with additional restrictions intended to make `EPS` files usable as a graphics file format. In other words, `EPS` files are more-or-less self-contained, reasonably predictable postscript

documents that describe an image or drawing and can be placed within another postscript document.

(http://en.wikipedia.org/wiki/PostScript - http://en.wikipedia.org/wiki/Encapsulated_PostScript)

route

Routing (or routeing) is the process of selecting paths in computer networking along which to send data or physical traffic. It directs forwarding, the passing of logically addressed packets from their source toward their ultimate destination through intermediary nodes; typically hardware devices called routers, bridges, gateways, firewalls, or switches. The term can also be used to refer to the physical traffic of data packets internally in a structured management system.

(http://en.wikipedia.org/wiki/Routing)

server farm

A server farm or server cluster is often a collection of computer servers usually maintained by an enterprise to accomplish server needs far beyond the capability of one machine. Another common use of server farms is for web hosting, which are sometimes referred to as *web farms*.

(http://en.wikipedia.org/wiki/Server_farm)

shell

In computing, a shell is a piece of software that provides an interface for users. Typically, the term refers to an operating system shell which provides access to the services of a kernel. However, the term is also applied very loosely to applications and may include any software that is "built around" a particular component, such as web browsers and email clients that are "shells" for HTML rendering engines. Command line shells provide a command line interface (CLI) to the operating system. The primary purpose of the shell is to invoke or "launch" another program and view the contents of directories.

(http://en.wikipedia.org/wiki/Shell_%28computing%29)

source code

Source code (commonly just source or code) is any sequence of statements and/or declarations written in some human-readable computer programming language. The source code which constitutes a program is usually held in one or more text files, sometimes stored in databases as stored procedures. A large collection of source code files may be organized into a directory tree, in which case it may also be known as a source tree. A computer program's source code is converted from human-readable form to some kind of computer-executable form. It can be converted into an executable file by a compiler, or executed on the fly from the human readable form with the aid of an interpreter.

http://en.wikipedia.org/wiki/Source_code

TeX - LaTeX

A scientific typesetting system allows anyone to produce high-quality document using a reasonable amount of effort, and it provides a system that gives the exact same results on all computers, now and in the future. TeX is considered by many to be the best way to typeset complex mathematical formulae. It is popular in the mathematics and physics communities, however, some have complaints about TeX such as unhelpful reference materials, off-putting jargon and cutesy terminology, inconsistent language design, and hostile user interface.

(http://en.wikipedia.org/wiki/TeX)

An example of the typesetting structure:

```
\documentclass{ some class }
\usepackage{ package }
\begin{document}
   \title{ title }
   \author{ author name }
   \institute{ institute }
\abstract{ abstract }
\maketitle
\section{ section title }
        section text
\begin{equation}
\bigl[ \pi(\eta,\vec x), \varphi(\eta,\vec y) \bigr] = - i\, \delta^3(\vec x-\vec
y) \qquad \pi = a^2\, \varphi' .
\end{equation}
\begin{figure}
\includegraphics[]{ figure.ps }
\end{figure}
\end{document}
```

verbose - verbosely

The term in computing draws from use to describe persons, writing and documents. The more a function or process is designed to feed out textual information on what it is doing, the more verbose. To process something verbosely (or not) is often an option on the command line, initiated by adding the relevant parameter to a processing command.

APPENDIX II

Notes on preparing data and convention for the use of notations and representing different types.

The preparation of these data was aimed at clarity and usability. I needed to describe and explain what was going on in a clear and succinct manner. I wanted to display accounts in a way that could help me work with much of the text as if it was natural language. Extracting single time-ordered threads of the interaction took some effort of cutting and pasting between documents and double checking for stray inputs that might have shifted position within the database. I also needed to make sure that actors were clearly distinguishable without revealing their identities.

All code, electronic exchanges and the pre-written messages are presented in the `Courier New` type face which best resembles the default system font on the `Unix/Linux` command line and in simple messaging devices. I do this for distinction and better clarity. Single names of compiled and packaged programmes are typed using this same type face, such as `Scientific Word`, `emacs` and `Adobe Illustrator`. When I refer to command-line scripts, however, configuration files, parameters or other source code and commands, I use single inverted commas, for example `` `lastmv' ``, `` `thebibliography' ``, `` `sendmail.cf' `` or `` `dvips -o -f file.dvi > file.ps' ``.

All digitally born accounts (interaction, exchanges and documents) which are quoted or displayed at length are left in their original font, spacing, margins and other formats, however, I replaced identities with descriptions or tags in bold face. Actual names of persons

are replaced with terms like **operator**, **submitter**, **colleague**, etc., and when more than one operator is involved, they are tagged as **AA**, **AB**, **AC** as they enter the scene. There is no correspondence throughout the text between these tags and actual persons. Submitters are **S01**, **S02**, etc., library system staff are **LS**, and so on. Other identities that are hidden concern individual packages, preprint identities and location paths on computers. For example if an exchange refers to a preprint with the identity `hep-th/0311021` it would instead read **`archive/papernum`** which basically holds the format for the identity. Sometimes the format is `hep-th/`**`papernum`**, or **`archive`**`/0311021`, depending on the context. Figures are typically renamed **`figure`**`.eps`, sometimes numbered if there are many of them in a list. A paper source file would similarly be renamed **`paperfile`**`.tex` and other files simply **`file`** with the relevant extension. Location paths are converted to, for instance, `home/`**`path/to/figure`**`.eps` and `http://`**`archive.domain/`**. This is done for the sake of clarity.

Once I had extracted single threads of exchange, each regarding some issue that involved multiple inputs, they needed some reformatting in order to display the time-ordered sequence. They have been provided with entry lines including the hour and minutes, who the message is from, to whom it is directed and that it is copied to the shared exchange:

`21:27` **AB**: ***to*** **S01** (*copy* ***to admins***)

The subject line is usually included as well because of the information value it holds, such as "**AA**:***to admins***. (*subject*: poor quality font (**archive/papernum**))". On one occasion I decided to use bold face and italics to distinguish one input from another because it provided much better readability.

I frequently cut accounts short with the […] notation to get at the core of some message or sequence and in the time-ordered sequences I usually omit signatures or initials but write into square brackets here and there for clarification and/or explanation: "Removed [the

temporary processing file] **file**.flg which must not be present". When the system is reporting, I also use bold face descriptions to replace identities and quotes from abstracts or titles: "astro-ph/**papernum**: contains exclamation mark !: **...quoting the abstract ?**". Bold face descriptive terms appear as well in my replication of a basic preparation of a TeX source file, that is to say in places were otherwise some text would be written.

Threaded view of message exchanges on two days in the administration (Appendices V) is explained to some degree in the actual message text which is more or less fully quoted. Outside contacts (submitters and all others) are tagged as **S01, S02**, **S03**, etc. The threads are numbered and they stretch back in time and into the future which is expressed with faded ink and indicated with a header of this and similar formats: f**rom 1 day earlier...**

The only audio recording quoted in this thesis was taped on an analogue dictaphone of rather poor quality. I am both the voice and the transcriber but there were uncertainties still about a number of words, mumbling, background noise, and so on. The main point of the exercise was to capture the work flow though, not every word. The recording is a monologue and monotonous for the most part, slow and trails off and back in at times which is expressed with a ... notation. Commas are typed here and there to clarify the pace, a breath or shifts in focus without an actual pause (relative to the pace). Where there are brief but untimed pauses, I have typed a period and started a new sentence and long untimed pauses are expressed with the (.) notation. I omitted cursing in Icelandic.

APPENDIX III

A list of subject archives in the system and a screenshot of the web submission interface (overleaf).

The subject archives are ordered by date of creation and including the abbreviations used for them.

Since 8/91	High Energy Physics – Theory
Since 2/92	High Energy Physics – Lattice
Since 2/92	Mathematics
Since 3/92	High Energy Physics – Phenomenology
Since 4/92	Astrophysics
Since 4/92	Condensed Matter
Since 7/92	General Relativity and Quantum Cosmology
Since 10/92	Nuclear Theory
Since 1/93	Nonlinear Sciences
Since 4/94	High Energy Physics – Experiment
Since 12/94	Nuclear Experiment
Since 9/96	Mathematical Physics
Since 10/96	Physics (general)
Since 9/98	Computing Research Repository (CoRR)
Since 9/03	Quantitative Biology
Since 4/07	Statistics

APPENDIX IV

Operator's voice at the terminal and a threaded view of the message exchange on the 21st February 2002.

Notes

- I am both the voice and the transcriber but there were uncertainties still about a number of words, mumbling, background noise, and so on. The recording is a monologue and monotonous for the most part, slow and trails off and back in at times which is expressed with a ... notation. Commas are typed here and there to clarify the pace, a breath or shifts in focus without an actual pause (relative to the pace). Where there are brief but untimed pauses, I have typed a period and started a new sentence and long untimed pauses are expressed with the (.) notation. I omitted spontaneous cursing in Icelandic.
- The threads are explained to some degree and are partially or fully quoted. Signatures are omitted and identifiable data replaced with bold faced descriptions as explained in Appendix II. Threads stretch back in time and into the future which is expressed with faded ink. Operators are **AA**, **AB**, **AC**, **AD** and **AE**, outside contacts (submitters, others) are **S01, S02**, **S03**, etc., and the system is **Sys**.

Operator's voice

Now, I usually start by looking at some of the email messages, here is a customer complaining about a blocked access. He is worried about not getting in as registered user to upload papers. (.) What I can do is to look him up and see if he is in the register file. Maybe he has his password wrong or he has been [his submitter status] kicked out, so I will communicate that to **AA** and **AB**. (.) We have something called archive documentation, where all the tasks that we do are listed, a lot of this… (.) I should just print this out to show what the types of commands we are actually doing here. I [am] going to look at the `standard email questions', something is here called `registration issues'. (.) It says that I should check the machine where this [request] is coming from and what the program is to run that [check], so I will try doing that. Now… (.) here is an example that the doc[ument] pages have not been updated or the program may not be available to the e-print [administrative] user

identity. We have had that before. I do not see any [relevant scripts], looking into the `bin' directory of the [main] server machine, other script that we… (.) it is not clear to me, which one of those could be an updated one to do this task. My guess it is not there at all. (.) So I am going to reply to this to [message exchange] and point out to **AB** that the doc[ument] page on access issues needs update. (.) I see there are plenty messages in here and I am running really late. It is already 13:24 and I haven't started reading through the metadata yet. There are 56 messages to look at. Here is someone requesting registration. Blah, blah, blah, I work for this research group and at the department of [research area]. Something in [location] and we have started to give standard replies to these kind of messages. We want to see sponsorship, authors need to find someone to sponsor their submissions here. (.) There are however cases where research groups do not have institutional accounts or web pages and one way to deal with this is to reply into [message exchange], hey guys, has someone heard of this group? Are they real people or what is going on here, (.) something of the kind… . (.) Somebody here [is] asking a detailed question about registration policy at the archive. Well, we are in the process of revising those so I have to answer that. This guy has been harassing us about registration policy… (.) office personnel could answer these types of letters. This is a matter of understanding of how they should be separated out from the main message mail body. (.) […*cursing in Icelandic*…]

Another mail… (.) **AB** is dealing… (.) someone who wants to harvest the archive for papers from Universities in Australia and dig deep into the system with text search… (.) some project they are doing there. **AB** is discouraging here saying to use the Google search mechanisms that will be activated in about a month or so. They will be indexing all of archive as we have become aware of here today… (.) bunch of Google correspondence. (.) Somebody from [a] net traffic research group, wants to track traffic from different sources. This is archive policy issue and we do not offer this kind of service to anybody. (.) What is else in here… (.) …hmmm (.) here is communication with moderators about a paper that was put on hold. It is crucial for archive administrators to be aware of moderator

suggestions and we always act upon their comments. (.) Hmm, is somebody from [physics institute] who wants to make connection [friendship] with archive. These messages come in weekly and here is email exchange with Google, which is quite exiting. These messages should maybe not be in a mail body for library administrators. (.) Here is someone thanking us for fixing his figure that only showed up as a white plane because the bounding box wasn't defined correctly for the image. **AA** dealt with this earlier. (.) Here is Mr. [author], we know him. Please tell me why my paper is rejected… (.) I can actually answer that saying that we have removed it upon a notice from our moderators. Moderator comments are not available, period. This could be standard reply by office personnel. (.) Here is is a letter of sponsorship that I have to take seriously, for a person trying to register in the past few days. This means I have to pipe a registration through manually. (.) Now this takes me again to the documentation pages, which are somewhere over here and so… (.) …right… . (.) On the command line I have to get myself down the directory tree on the server and find a bad registration requests in a special folder respectively. Everything goes into the system somewhere, and rejections of all sorts flow into various directions where they can be looked at internally and manually if you know where to find them somewhere down the tree. Now there was **AA** coming in, replying on [about] this other person from the [location]. (.) Let him in, and now I have two registrations to pipe. Here is one of these dilemma questions. Someone reports that they are having a problem adding a journal reference. Now I can choose between referring him to our help page and not bother even checking the logs… (.) deeply buried. I can check the logs and communicate what went wrong, or since he included the reference he planned to add, do it for him. This is an interesting case study. I should not do it for him at least. (.) I have to start reading abstracts. Usually I start whizzing through the replacements [new versions], those are easier… (.) still 58 of them as of this moment now, 14:08. (.) We read through replacement metadata, primarily to make sure there is nothing odd about the size of the new paper version and we scroll down to see the list of uploaded paper files, if anything is wrong with that list, and if there is a problem with the internal postscript [generation] then that gets flagged at the top of the metadata page. I do not have to worry about any

other issues on paper replaces, such as the author field. If the new version is much smaller, something might be missing. Another scenario would be that they [authors] are withdrawing a paper and that works like a replacement but comes in without any paper files so the size [indicator] is down to zero. (.) If it is a much bigger submission some new figures may have been added. (.) Reading through replace metadata shouldn't take more than a few minutes, 10-15, given that nothing comes up. (.) Here I have a bad postscript generation [warning]. I am going to walk through that in a minute. This is being flagged by the system because it can't generate postscript [on the fly]. (.) Ideally with this type of warning, first [task] is to grab the paper identifier and see what it [the paper] looks like on the archive web [site]. First see if it generates a PDF [from the postscript file]. If it doesn't something is obviously wrong. It does. (.) …hmm interesting (.) …hmm… (.) …ok, the last page is an image, a model, it is a big image file so the ghostscript generation times out and flags this warning. Postscript generation is [working] alright, it simply takes ghostscript quite a while to draw that image. PDF is much faster. (.) And now I am going to make sure that the postscript actually works… [look at it in a postscript viewer] (.) it does. (.) Here is an example of author replacing an older paper, sending in both a PDF and a PS [postscript] file. This is forbidden, period. (.) Sometimes authors stick all their images into a separate PDF [file] or they are sending us two versions, one in each format. Whatever, this gets kicked out and this is a little more complex than regular submission kick-out because this is a roll-back to a previous version. And now I am back to the documentation, I have this in my head by now. (.) First I check what is available as primary paper files for this paper-identifier [the new replaced version] and now I have to crawl into the directory structure, to where the originals reside [older versions], and move the previous version back overwriting the new incoming version, and make sure that any extra additional files are manually deleted… (.) and send the author a letter. (.) I see a replacement size decreasing down to one tenth the original submission… (.) comments say it is only four pages and no figures but I am still going to look at its web-appearance. It seems to be a whole paper, nothing is missing but this is only something the eye can catch and is not flagged by the system at all.

(.) Another example is of authors turning their names the wrong way, last names first. We have discussed the design for a control here, but I think we should for now make it a habit of writing the authors a note about this and that we are unhappy and that it is a shame that they are not taking care of their metadata. We have to manually edit the data in a command line editor. (.) It is already 14:20… (.) this is scary. I am still not reading any abstracts [metadata] for new submissions. (.) Do that now. I try to do this surgically. I skim the titles, I read author names and their affiliations, make sure they will parse right, no open parenthesis, no last names first, also check for special characters, that they are expressed right so they display right. I also skim the abstract for anything unusual and scroll down to the ‘contains' section where the list of uploaded paper files is and make sure my eyes catch wrong file types together or if the `comments' sections says there are 6 images and it turns out there really are none. (.) Gets kicked out, something of the sort. (.) Here is an example of last names first. This system will [should] flag this as a problem at 15:15 [pre-freeze time] and it needs manual fixing. I might just as well do that right now. I grab the paper identifier and switch over to the terminal into the server machine and type ‘abs' and the identifier, which opens the metadata file in a text editor and I manually switch the last names and the initials around. (.) Here is another example of a BAD PS generation warning and the author has twice tried to replace his upload to fix this. (.) I like to wait with these problems until late in the day, to give authors a chance to fix this themselves, which sometimes is the case. (.) PDF and PS only submissions are flagged by the system at 15:15 but they need to be manually looked at and removed if it is a file that is copyright of a journal… (.) if it is a file generated from TEX and/or using type3 bitmap fonts. (.) Sometimes it is good to be done with that before 15:15… (.) especially on a busy day. (.) Here is an example of a Bad PS warning that has obviously been fixed by the author, no more bother with that. (.) Here is still another example of a BAD PS warning that has obviously been fixed by the author, no more bother with that. (.) A few metadata problems… (.) here is an example of a star [*] being used in author field. We allow for the use of numbers in parenthesis that are associated with listed institutions by number. Again take the paper

identifier onto the server machine to manually change this. (.) We also remove any postal code, zip code… [from institutional affiliations] (.) no postal addresses, only where that inst… . (.) One little ridiculous thing here. I am doing cosmetics on an abstract here with unnecessary dollar [math] signs all over, where it really isn't needed. And these expressions for italics and stuff in TEX… (.) it looks stupid, so… (.) …this is not something I should be doing, but what do you not do for the archive to make it look good… (.) talk about dedication… (.) this also needs a cosmetic fix, typical… (.) same author. (.) Here is an interesting [metadata] case. In `report number' field, there is a URL, which cannot be there and a mention of a journal reference when there is no journal reference. This is junk and I move manually the URL into the `comments' field and cut the rest out. (.) It [this] is another something that the system flags but still needs to be manually done, its tildes [~] where there should be a white space. Sometimes it [a tilde instead of white space] turns out to be the case, sometimes not. (.) Oops, an advertisement for a book. It is an appreciation and a book review that is meant to be published somewhere. archive does not host that at all. It is not the type of presentational material that… [we] do serve. Now I will remove a submission from the system. (.) Trying [the command] `lastmv paper-identifier', is no go because this is the last one to come into this archive, so I have to pop the last paper number off that archive folder [command `lastrm' instead]… (.) and send the person a letter, explaining this. (.) 15:24 I still have 20, 21, 22, 23, 24, 25 metadata files to look at, but it is pre-freeze time [already] so I am going to have to… (.) there have been some BAD PS [warnings] and I like to wait with these until last minute, because authors often are responsible enough to fix them, but sadly often they are not and they [the submissions] have to be kicked out unless we can fix them somehow. I have to whiz through this [the rest of the metadata] right now and then get to the checking of BAD PS [warnings], and look at the PDF and PS only [submissions] and I even suspect there is a html file [submission] around here today. (.) Until now I have only been reading abstracts [metadata] since I stopped looking at the mail, done some manual fixes of it… (.) here is someone who forgot to put [a] comma between names, they [authors] don't even look at this

stuff to fix it. The author names are not searchable as it is. So… (.) the system again flags a situation like that, but that is not enough. It still needs manual fixing. It is good that it flags these things, since the eye may not catch it, but I haven't even started looking at what the system is flagging today. I just came across this one so I am fixing it. (.) Three BAD PS to look at. (.) First step is to look it up on the web and see what it looks like, try to generate a PDF on the fly. That is not happing so we are sure to have a problem, so I check the postscript generation and see what warnings it will give me. That is a long paper. All right, it is a figure problem, Ok… (.) this looks like binary junk or… (.) what I have to do is to move over to a terminal on my own machine where I have seamless access to the server for the purpose of hacking [copying] trouble papers over onto my machine to run tests on them [with the script `pscheck']. An `00README.XXX' case [additional directive for the auto-compiler]… (.) maybe there is a [document] file for this. (.) I re-compress the package and restore it on the server machine and now I can check again on the web if the paper is displaying properly. Ok, it is coming through, this is good and the PDF is working fine as well… (.) good. (.) Two more BAD PS [warnings] to go. (.) This is bizarre, because the images are not `inline' in the document and I'm reading the message to the author. Ah, the images are all bad, every single one of the, 7 of them. I have to look at that. (.) A phone call [came] from **AB**, about the fellow who received a message when his submission [removed] was removed because it was mixed file formats, PDF/PS. (.) I sent him an obvious message, he should have understood that. He is just going to get a way with this. Some people are just odd. (.) Anyway, I didn't even look at the content of the files, just kicked the submission out. It [this] is a replacement. Well up his. Wait a second, there is only one file here now. He submitted a PS only this time. But he can't do [submit] it anyway because it is copyright of a journal. OK… (.) I also mentioned to **AB** that I wrote a script to send in and count abstracts for all [archive] lists and to grep [search] for journal titles in all of them and we discussed a bit some precautions to keep in mind. (.) So, this was an interesting example, the author submits again only one of the two files and it turns out to be copyright of a journal. That… [is a] version removal and that is

a drag. (.) I should move quickly through the PDFs because it looks like I am going to run out of time, it is 15:43. We can have the scans through PDFs [checking through a list of them], which is an easy job, somehow flagged to office personnel… (.) they may have to be removed. What is problematic about them is that the removal process at this point is not very straightforward for a GUI [graphical/web user interface, (also WUI)] oriented computer user. (.) We could see that better automated with GUI… (.) button stuff. (.) We open all PDFs and PS only submissions, check font info for the PDFs, see how it is produced. If it looks like it is produced from TEX it is removed. Here is one obviously, either produced from TEX or no… (.) it is a bitmap, a scan. I wonder if it is copyrighted… (.) …no, but goes. So that is a `lastmv' process. The submission taking the place of this [removed] one, needs to be checked briefly [also], just to make sure [that] that file doesn’t also need removal later. No need to be shuffling submission around forever and sending out repeated notices to authors. (.) Out of 6 PDFs, two have to be removed, because of type3 bitmap font and an image scan. I am checking quickly on a single PS the system flags, that is certainly the new version that I sill need to remove again, so… (.) so, what other things are being flagged here by the system at 15:15?… (.) lets see, there is possible third party submission… (.) and we check those by looking them up on the web. If there is no comment field… (.) we check those too in the same way. It is flagging unnecessary formatting in metadata… (.) a few of those, on a busy day, there can be many of those. Something listed by author in comments field as being on request. We check [those] because we like to make sure that all the images are always accompanied with the papers, not just available on request from somewhere else. If anything is missing we kick the submission out. (.) Here is a good example of how the system flags tildes that might just be occupying white spaces, but the system is incapable of correcting this without human approval, as in this case, where the tildes actually should be there. I am wondering at this point, how much [trouble] will be out of the picture when all incoming metadata will be converted to [the character set] `charset-UTF-8'. (.) Something may not be human readable… (.) citations. Hmm yeah, a whole bunch. The system flags it but you still have to go and manually fix it.

Ok, and that is not something I will be doing right now, because it is not quite as urgent, I will not be doing that before 16:00, only right after. (.) Aright, ok here is that BAD PS [warning] with all the figures bad. I am going to look at that quick. I will have to hack [copy] the whole submission onto my machine. It is probably a Macintosh user, generating and compressing figures that are full of binary junk, which I end up cutting out manually in a `vi' [command line] text editor. Yup, all have some junk in them. It is good to know that this can be fixed, but this type of scenario is administratively bad because it is unclear if it is entirely author responsibility to deliver good figures, or if I am acting a part of a common sense support structure in the archive environment. (.) Looking it up on the web just to make sure it is displaying fine now. (.) 15:59, and I still have one BAD PS [warning] to look at. (.) Here I am checking for moderator comments in the busybody mail just in case and this is important to do before [the] 16:00 [freeze]. Here is a suggestion that a paper submitted to `nucl-th' be submitted to `cond-mat' and since we always do what they say I will have to route. Now, what I really do is that I basically remove the paper, as previously described [with `lastmv'], jamm, no kidding, and then I move myself into a sub-directory `TMP' and `reput' the paper into the system from there, pipe it into the other news group… (.) `cond-mat'… (.) …and now [back to] looking at one of the BAD PS [warnings]… (.) it looks like a figure is screwing up here. I am going to hack [copy] this [submission] and see what I find. (.) All… (.) …right. This a another major junk submission. A few image files are junk for the most part. Authors should be notified about their files being like that… (.) it seems like they are not acting upon the warnings given to them. (.) 16:08, I hope I am not forgetting anything major important but now I still have to edit manually some of the metadata. This is a relatively easy day, only a 112 new submissions and about 62 replaced older submissions. (.) I am getting back to the busybody mail. Here is a letter from our infamous submitter who tried to replace an old version of his paper twice today, `…ah, I see the problem, well I don’t have a version, don’t even have the figures…' (.) Sorry, mister, the stuff you are sending in is copyrighted…' (.) no go. (.) Here is an interesting letter, someone in `astro-ph', worried that the figures of his manuscript are just dancing

away. We are really not equipped to deal with consults on this sort and I usually flag the messages to **AC**. However, let me hack [copy] the submission just to see what I find and here I can see that his document definition is for a A4 size paper instead of US letter size, so I am sending him a note telling him that… (.) **AC** [just] came back to me about my reply to this author, and about another TeX case as well. I was guessing the wrong things. (.) Here is someone wondering about a paper that has obviously been removed by somebody, it is a PDF and PS and TeX mixture and I cannot find the paper-identifier in any of the mail, so I assume it is our wild card. We have one of those on the loose in the system every now and then. We call it **AA**. (.) […*cursing in Icelandic*…]

Here is someone writing us because of registration… (.) wants to post a discovery and has no backup and what to do… (.) sorry we never register unaffiliated authors. That's it. (.) Finally, we do have a couple of interesting examples of my-too-large-submission inquiries. (.) Hacking does not only enable me to grab new submissions that come in, but also to grab any rejected submission by calling its rejection id. Everything that comes into the system goes into some place in the directory structure. (.) Now here is a rejection because of size, where author is simply thoughtless and generating needlessly big images. My way of dealing with this is to point out a couple of figures in their package and then there is a standard letter that encourages authors to think about the archive not as a venue for the exchange of large images and datasets, they could store low quality copies with archive and refer readers to their own website somewhere holding full resolution images. We should be rethinking the way the detection takes place. The system is rejecting total size, should maybe be checking single file sizes [better]. (.) Here is another submitter with the same problem so I am going to hack that submission. Not only has this person… (.) the source files of this submission are sitting deep in a directory tree, one, two, three, four, five, six… (.) six nodes down the tree, which is very unusual. We like to see the source files wrapped up at the top level. (.) Hmmm… (.) Anyway. We point that out to the author and the submission includes enormous figures so we send her that same standard letter of encouragement. (.)

Threaded view, 31 message threads

- - - 1 - - -

from 5 days earlier...

08:45 **S01**: ***to admins*** (*subject*: problem with submitting tex-file managed with CVS)

S01 is trying to get in touch with the admins about a problem detected with the TeX auto-compiler.

“Today I had problems submitting to the archive. I have solved the problem but others may experience the same so I thought it could be worth telling about it. The problem was that my .tex file was taken to be a postscript file by the automatic system.”

The reason as **S01** explains it, is the use of the Concurrent Versions System CVS which uses a control line for detecting version number of documents, here **S01**'s TeX file, which begins,

“%! $Id: ms.tex,v 1.18 […].

Postscript files start with %!PS but apparently the automatic system only checks the first two characters.”

from 2 days earlier...

12:03 **AD:** ***to admins*** (*subject*: Re: problem with submitting tex-file managed with CVS)

“??”

16:44 **AA:** ***to* S01** (*copy* ***to admins*** (*subject*: Re: problem with submitting tex-file managed with CVS))

“Thanks for your comment. Sadly we can[not] rely on anything more than "%!", there are so many programs that write non-standard PS! This is a heuristic that works most of the time.”

...on the day...

10:20 **AC:** ***to admins*** (*subject*: Re: problem with submitting tex-file managed with CVS)

“it's conceivable we should change [the script] fileguess to detect %! $Id ... and exclude that as PS, if this is standard for CVS of tex files. our CVS on perl files uses instead # [CVS: $Id: get.ph,v 1.4 …]”

AC also mentions having,

“modified mailbin/comments.ph since i got tired of my messages to www-admin being bounced, needs check-in”

15:49 **AA:** ***to admins*** (*subject*: Re: problem with submitting tex-file managed with CVS)

“I don't think this is standard.”

Then **AA** quotes command-line argument and parameters for finding examples of controlled TeX files, and what the machine reports back.

“If I look at my cvs controlled tex files I find:

[**AA**]@[**computer**]>find ~/docs -name "*.tex" -exec grep '$Id' {} \;
[\verb|$Id: … $|]
% [CVS: $Id: … $]
[…]
% CVS $Id: … $

{\tt \verb|$Id: … $|}
\verb|$Id: … $|\\
[…]”

The system reports back at least 5 variations on how control strings for CVS can look like.

“I just happen to put [CVS: $Id: ...$] Only the $Id: ..$ part is actually used by CVS”

On **AC**'s request of checking into his alteration of “comments.ph”, **AA** replies, “done”

- - - 2 - - -

from 3 days earlier...

13:04 **S02**: ***to admins*** (*subject*: archive and […])

This is from an academic publisher who manages an abstracts search service, says that,

“[…] we like to keep track of how much traffic we drive to the different sources. As only the receiving side can provide us with the number of referrals generated (regretfully we can not generate these figures ourselves), we would be very very happy if you would be willing to provide us with these figures. […]”

from 1 day earlier...

20:24 **AA**: ***to* S02** (*copy* ***to admins*** (*subject*: Re: archive and […]))

“I'm afraid we don't offer this sort of information.”

on the day...

10:17 **S02**: ***to admins*** (*subject*: Re: archive and […])

“Okay; I understand; just thought I could try!”

- - - 3 - - -

from 1 day earlier...

06:24 **S03**: ***to admins*** (*subject*: 0202224)

S03 claims to have a problem with a figure in a paper downloaded from the archive website.

“If I create a pdf file, I can read the paper no problem but the figure doesn't come out. The ps version of the figure just produces a postscript error. I'm using Acrobat 4.0. All suggestions gratefully received!”

15:57 **AA**: ***to* S03** (*copy* ***to admins*** (*subject*: Re: 0202224))

“Figure had messed up bounding box. Fixed (and included with the text). Change will propagate to mirrors overnight.”

on the day...

03:56 **S03**: ***to admins*** (*subject*: Re: 0202224)

“Thankyou!”

- - - 4 - - -

from 1 day earlier...

15:55 **AA:** ***to admins*** (*subject*: Re: robopage: web 500 15:11)

“Hmmm... I clicked on PDF for user supplied PS before public.”

...meaning **AA** attempted to invoke the online PDF conversion for a new submission already built as postscript but still not been made available online, except to the submitter.

“Probably something to do with group priv?”

`priv' is a user group name and identity for all submitters to access and test their submissions online while submissions are still access restricted to the public. **AA** quotes two logs, the web server access log showing **AA** making a PDF request with `internal server error' results, and the web server error log reporting the client (**AA**´s machine) invoking the request,

“[p]remature end of script […]”

...indicating that group identity `priv' for the package is failing.

on the day...

10:56 **AC:** ***to admins*** (*subject*: **Re: robopage: web 500 15:11**)

“did you look at all the error_log entries for this one? problem is at if (!unlink($file)) { [an if-clause] in [header file]”

AC quotes log reporting from three header files, warning first that the metadata for the submission **AA** calls for is `private'. Then since **AA** makes the call as a member of a different user group (although of more privileges) than `priv', the log reports that authorisation dependency is insecure in [a script that un-links the postscript file for PDF conversion] `unlink', while running setgid [setting group identity] for it. Then **AC** also quotes the same error from the web server error log,

“[p]remature end of script …”.

- - - 5 - - -

from 1 day earlier...

16:53 **AE:** ***to admins*** (*subject*: leftover issues (02/20/02))

AE reports to admins about leftover issues on the previous day, including,

(1) issue about a withdrawal notice while on the website the package,

“still offers pdf/ps links that are invalid. not sure on this ???”

(2) five crosslistings of one submission,

“hep-th/**papernum**”

(3) “question on large submission via email.”

...that has not been dealt with

(4) an issue with an abstract containing footnote,

“i removed slur at end of abstract. not sure how you want to handle footnote.”

(5) a “registration exception request.”

...that has not been dealt with.

20:20 **AA**: ***to admins*** (*subject*: Re: leftover issues (02/20/02))

(1) “This is an old problem. Fixed.”

(2) “actually only 4 [cross listings] since math-ph and math.MP are one doesn't seem too bad and **AC** hasn't objected”

(3) “[…] why couldn't you answer? standard reply is that we must get the rejection id to look at oversize submissions. Replied.”

(4) “i commented out parts of the tex, replied to [submitter]”

(5) “piped though with the university address he claims not to be able to send from.”

on the day...

11:01 **AC**: ***to admins*** (*subject*: Re: leftover issues (02/20/02))

(2) “i'm not tracking the crosses from hep-th, only to hep-th”

(3) “except this could run into problems since after all the back and forth he might receive some message about how we need sponsorship due to new guidelines”

- - - 6 - - -

from 1 day earlier...

12:27 **S04**: ***to admins*** (*subject*: RE: FW: register)

“I'm trying to register at archive.domain and received the message below, presumably because my dns [domain name server] does not have ".edu" as an extension. I'm a Science Fellow at ... […] Please let me know if you require further information to process my registration.”

13:28 **AD**: ***to* S04** (*copy* ***to admins*** (*subject*: RE: FW: register)

“Due to numerous problems, we are in the process of revising our policies with respect to unaffiliated researchers. If you have no recognized academic affiliation, then you must find someone who does, and with expertise in the relevant subject matter, to sponsor your submissions here.”

17:58 **S04**: ***to admins*** (*subject*: RE: FW: register)

“[**Professor**] of the [**university**] is trying to get a response to you...he's at [**email address**]. His e-mails keep getting rejected. By the way, does archive.domain reject registration by staff at IBM, Bell Labs, Xerox, GE, LANL, ANL, ORNL...? They're not dot-edu's either.”

18:32 **AD**: ***to* S04** (*copy* ***to admins*** (*subject*: RE: FW: register)

“[**Professor**] can write his letter of sponsorship directly to: www-admin@archive.domain

We have registered authors from labs like MSN and more. The registration of these individuals is treated as 'specical cases'”

18:35 **S04**: ***to admins*** (*subject*: RE: FW: register)

“Is there a detailed statement regarding institutional registration policy at the archive website? I couldn't find one.”

on the day...

13:39 **AD**: ***to* S04** (*copy* ***to admins*** (*subject*: RE: FW: register)

"We are in the process of revising our policies. They will be made public as soon as the work is completed here at Cornell University."

14:11 **S04**: ***to admins*** (*subject*: RE: FW: register)

"Have you received the sponsorship e-mail from [**professor**]? He just re-sent it."

- - - 7 - - -

from 1 day earlier...

20:48 **S05**: ***to admins*** (*subject*: blocked access)

"I was trying to access from [**machine name**] [**IP address**] and found out my access blocked! […] Would like my problem to be solved when possible."

on the day...

13:32 **AD**: ***to admins*** (*subject*: ! ADMIN Re: blocked access)

"Document page on access issues need an update."

AD quotes the document page,

"Take the machine name first (also look for the IP address the same way). Go to [**main server**] and type graccess.pl machine name."

16:12 **AA**, ***to* S05** (*copy* ***to admins*** (*subject*: Re: blocked access))

"Access has been unblocked but please note that our automatic robot detection system will replace the block if an `attack' is detected. This can also be triggered by `accelerator' or pre-fetching software. You must remove or disable such software before accessing our site. Why? See: http://archive.domain/RobotsBeware.html"

16:13 **AA**: ***to admins*** (*subject*: Re: ! ADMIN Re: blocked access)

"was this from [**local document pages**]? I have updated so that hopefully it makes a little more sense, comments welcomed."

16:41 **AD**: ***to admins*** (*subject*: Re: ! ADMIN Re: blocked access)

AD quotes a command and machine response to report that the script `graccess.pl' is not available to the standard administrator user identity,

"[e-prints@**machine**]$ graccess.pl [**IP address**]
bash: graccess.pl: command not found

??"

17:32 **AA**: ***to admins*** (*subject*: Re: ! ADMIN Re: blocked access)

"Worked when logged in as me ;-) Okay moved these to ~/bin on both [**main server**] and [**backup server**]. Should work for you now. Thanks for telling me."

- - - 8 - - -

11:45 **AC**: ***to admins*** (*subject*: Re: archive and Google)

AC copies an exchange with technical personnel at Google about parameters for Googlebot (the robot agents from Google) to `crawl' the system in order to harvest full text contents of the collection. They are communicating upcoming `testcrawl', starting with the computer science (cs) archive. **AC** asks if Google robots,

"support the Allow: syntax in the robots.txt [standard robots permission configuration file]"

...claiming that this file has been altered to accommodate for Googlebot, while strictly denying access to all other robot agents.

"User-agent: *
Disallow: /*
[…]

User-agent: Googlebot
Allow: /archive/cs
Allow: /list/cs
Allow: /abs/cs
Allow: /pdf/cs
Allow: /html/cs
[…]"

12:29 **AC**: ***to admins*** (*subject*: Re: archive and Google)

AC copies a letter from Google personnel saying that,

"Yes, we do support "Allow""

...and **AC** replies with confirmation of the contents of `robots.txt'.

a follow-up two and three days later...

[follow-up omitted]

- - - 9 - - -

06:30 **S06**: ***to admins*** (*subject*: rejection, bis)

S06 has received a rejection response from the system and now wants to know,

"[…]why has my paper ``[**title**]'' been rejected? Please, see: [**quotes auto-rejection response from an email upload attempt**]"

13:43 **AD**: ***to* S06** (*copy* ***to admins*** (*subject*: Re: rejection, bis))

"It was removed upon a notice by our moderators. […]"

- - - 10 - - -

09:36 **S07**: ***to admins*** (*subject*: Fwd: sponsor for Dr. [**S04**])

"[…] Attached is a letter of sponsorship for [**S04**] that I have been trying to send to you. Please acknowledge your receipt of this email to me and [**S04**]. Thank you. [**signature and sponsorship letter**]"

- - - 11 - - -

10:04 **S08**: ***to admins*** (*subject*: request for registration)

S08 is requesting to be registered with the system because it has auto-rejected the,

“e-mail address [which] is that of the […] research group at the Department of Physics [… at a university].”

...an academic/institutional address is not recognised by the system.

11:04 **AC**: ***to admins*** (*subject*: Fwd: request for registration)

“/**path**/**to**/**rejected**/[**registration-data**] should be piped through […]”

13:36 **AD**: ***to admins*** (*subject*: ADMIN! Re: request for registration)

“Has anyone heard of this [research] group ?”

13:56 **AD**: ***to admins*** (*subject*: Re: Fwd: request for registration)

“I'll do it […]”

- - - 12 - - -

10:44 **S09**: ***to admins*** (*subject*: (TRUNCATED) **archive**/**papernum** v2: Problem with .pdf format)

“[…] We replaced our **archive**/**papernum** yesterday with a revised version. There is a problem with downloading the PDF version:

PDF unavailable... [...**quoting further the message displayed on the archive website**]”

15:36 **AA**: ***to* S09** (*copy* ***to admins*** (*subject*: Re: (TRUNCATED) **archive**/**papernum** v2: Problem with .pdf format)

“PDF now available.”

- - - 13 - - -

11:24 **S10**: ***to admins*** (*subject*: journal reference)

“[…] My today's attempt in adding a journal reference to my existing work
(**archive/papernum**) was rejected. I would like to add [**journal reference**].”

15:43 **AA**: ***to* S19** (*copy* ***to admins*** (*subject*: Re: journal reference)

“You must include the year”

- - - 14 - - -

13:10 **S11**: ***to admins*** (*subject*: ADV: Quick Question)

“[…] I hope I'm contacting the right person here. My name is […] and I

am inquiring to see if you were looking for financing on any equipment […]”

[ignored]

- - - 15 - - -

13:37 **Sys**: ***to admins*** (*subject*: **archive** hput **paperfile**.tex **papernum**.gz 6kb from **submitter@address**)

Sys routes a copy of metadata for a submission that was uploaded via a mail protocol (`hput') and taking it as a message because it detects the word `comment' in the bibliographic data.

[ignored]

- - - 16 - - -

13:53 **Sys**: ***to admins*** (*subject*: Re: register from **S04@address**)

Sys reports that a,

“registration [of **S04**] has been piped through manually”

...on the `back end'.

- - - 17 - - -

13:58 **Sys**: ***to admins*** (*subject*: Re: register from **S08@address**)

Sys reports that a,

“registration [of **S08**] has been piped through manually”

...on the `back end'.

- - - 18 - - -

14:25 **AD**: ***to* S12** (*copy* ***to admins*** (*subject*: new version removed (was **archive/papernum**))

This is a standard note to **S12** about removing a package (here a replace of an older version) from the pending queue on the grounds that,

“[w]e do not accept mixed paper file fomats PDF/PS. TeX is our preferred format, w/ figure inclusions. Otherwise we do accept PS or PDF only submission, given that they are not produced from TeX. Please try resubmit see our help pages at: http://archive.domain/help/submit”

14:50 **S12**: ***to admins*** (*subject*: Re: new version removed (was **archive/papernum**))

“We published the paper in PRL [Physical Review Letters]. They won't release the source for the final version so I am left with only the PDF/PS. The .ps that people get from archive is out of date and I would like to replace it. What should I do?”

[ignored]

- - - 19 - - -

14:50 **AC**: ***to admins*** (*subject*: Re: did you see this?)

AC copies an exchange with a moderator for the condensed matter archive, about, what the moderators comments on as an,

“amusing attack on prl [Physics Review Letters] from a computer scientist...”

AC responds,

“ooh, i like it. tks for pointing out.”

...and directs the message to the attention of ***admin***,

“(admin note: definitely should remain on cond-mat.stat-mech)”

...or in the statistical mechanics subject class for condensed matter.

- - - 20 - - -

15:15 **Sys**: ***to admins*** (*subject*: prefreeze check (45 min to go))

The **Sys** prefreeze report reminds administrators to,

“RUN /tmp/autofixes”
...to do certain specified,
“[m]anual checks”
...and to check the,
“[c]ross [l]istings”
...of submissions over the day.

15:55 **Sys**: ***to admins*** (*subject*: prefreeze check (45 min to go)(titles))

Sys reports a list of titles for all new and replaced submissions till that point in time.

- - - 21 - - -

15:20 **AD**: ***to* S13** (*copy* ***to admins*** (*subject*: submission removed (was **archive/papernum**)))

A note to **S13** about removing a package from the pending queue on the grounds that,
“[t]he archive does not accept book reviews or any kind of advertisement for book publications.”

- - - 22 - - -

15:47 **AD**: ***to* S14** (*copy* ***to admins*** (*subject*: submission removed (was **archive/papernum**)))

A note to **S14** about removing a package from the pending queue on the grounds that,
“[w]e do not accept low quality PDF using bitmap fonts. Your submission looks as if it was produced from tex. If that is the case then you must submit the tex source. See:
http://archive.domain/help/faq/whytex.html”

- - - 23 - - -

15:55 **Sys**: ***to admins*** (*subject*: final freeze status (5 minutes to go!))

Sys reports what the freeze status is, and reminds administrators to,

“RUN /tmp/autofixes”
...to do certain specified,
“[m]anual checks”
...and to check the,
“[c]ross [l]istings”
...of submissions over the day.

15:55 **Sys**: ***to admins*** (*subject*: final freeze status (5 minutes to go!)(titles))

Sys reports a list of titles for all new and replaced submissions till that point in time.

- - - 24 - - -

16:00 **Sys**: ***to admins*** (*subject*: daily 16:00 freeze done)

Sys lists a summary for each archive: the number of new submissions, new cross listings, the number of replacements of submissions and replacements of cross listings,. Finally the total over the day is summed up,

“112 new papers, 53 new crosses, 51 replaces, 30 replaced crosses”

16:00 **Sys**: ***to admins*** (*subject*: journal-refs)

Sys reports all journal references added to submissions on this day.

- - - 25 - - -

16:00 **Sys**: ***to admins*** (*subject*: **archive/papernum** failed to produce postscript)

Sys reports that it fails to auto-compile a source package of a submission from January 1996.

“Use of uninitialized value in string eq at **/path/to/module** line 245.”

- - - 26 - - -

16:18 **AD**: ***to* S15** (*copy* ***to admins*** (*subject*: new version removed (**archive/papernum**)))

AD has removed a package (here a replacement of an older version) from the pending queue on the grounds that,

“[y]ou do not have the right to submit PDF/PS that is copyright of a journal. You may instead submit your own version. Be sure to submit tex source if you used tex.”

17:00 **S15**: ***to admins*** (*subject*: new version removed (**archive/papernum**))

“I see the problem. Well, I don't have a version. I don't even have the figures. So if it violates a copyright, please just leave the unrevised version. Thank you.”

- - - 27 - - -

16:58 **AD**: ***to* S16** (*copy* ***to admins*** (*subject*: your submission (**archive**/**papernum**))

AD sends a note to **S16** about a submission that was cleared on the `back end' to,

“point out that it is the responsibility of submitters to make sure that paper submissions comply with archive rules and that paper files display correctly. […] Four of your figures had enormous amount of binary junk in them. This junk was removed manually on our end today. Please keep a note of this in the future.”

continues the next day...
05:02 **S16**: ***to admins*** (*subject*: Re: your submission (**archive**/**papernum**))

“sorry. I was aware of problems but could not solve them at this time and was waiting for local help. Should I still replace the old version by a better one, or did you repair the defects so that no further action is neccessary? […]”

08:21 **AD**: ***to* S16** (*copy* ***to admins*** (*subject*: Re: your submission (**archive**/**papernum**))

“The problem is solved. This is a tricky situation because admins at the archive learn much from solving some of the incoming problems and fortunately we can sometimes communicate this back to users for everyone's benefit. The case of binary junk is difficult because it is really not up to us to teach users (esp. mac-users) to generate ps-figures and compress them correctly. […]”

- - - 28 - - -

17:01 **AD**: ***to* S17** (*copy* ***to admins*** (*subject*: your submission (**archive**/**papernum**))

AD sends a note to **S17** to,

“point out that it is the responsibility of submitters to make sure that paper submissions comply with archive rules and that paper files display correctly. […] All your figures had a line of *junk* at the end of them. This junk was removed manually on our end today. […]”

- - - 29 - - -

17:01 **S18**: ***to admins*** (*subject*: aps.archive.domain rebuild)

S18 from the American Physical Society (APS) reports a new set up and configuration of the archive mirror at APS. **S18** directs his words to **AA** and **AB** to,

“…please log in and take a look to see what's missing, and I should be able to get things fixed up.”

S18 also apologises for the delay, citing among other things,

“increased physical security”
temporarily barring physical access to
“some extra kernel configuration work (the network card in particular, which was one reason physical access [to the machine] was necessary).”

17:43 **AA**: ***to* S18** (*copy* ***to admins*** (*subject*: Re: aps.archive.domain rebuild)

"I can get to it with ssh [secure shell], but I can't log in. Please add the following ssh2 key: [**key**] […]"

continues the next day...

15:50 **S18**: ***to admins*** (*subject*: aps.archive.domain rebuild)

"Ok, try again... looks like it should work. By the way I can give you the password for the **administrators** user [account] if you like, or reset it again. Call me if you want [**phone number**] […]."

18:55 **AA**: ***to* S18** (*copy* ***to admins*** (*subject*: Re: aps.archive.domain rebuild)

"…The key works. However, it appears that I have [sitting on the aps mirror] a .login or .bashrc [configuration files] set up to run [the shell program] tcsh which isn't available: […]
bash: exec: tcsh: not found
Connection to **server** closed.

this dumps me straight back out. […]"

20:51 **S18**: ***to admins*** (*subject*: aps.archive.domain rebuild)

"Yup, [the shell] tcsh isn't available; let me know if you need it. Logins had worked for me - there seems to have been a line in there specific to you, [**AA**]! I commented it out anyway."

- - - 30 - - -

17:20 **S19**: ***to admins*** (*subject*: Re: **temporary identifier**)

S19 has a problem with the TeX auto-compiler because of errors that are visible on his end.

"[…] When I process the paper locally, I get the same error messages about a missing "}", but the errors are not fatal… […]"

...meaning that he presses the `enter' key to continue the build successfully on his end.

17:36 **AA**: ***to* S19** (*copy* ***to admins*** (*subject*: Re: **temporary identifier**))

"[…] You must supply tex that compiles without error (there is no automatic way to tell if errors should be ignored [or not] so our scripts don't). […]"

- - - 31 - - -

17:23 **S20**: ***to admins*** (*subject*: RE: (REJECTED) What program should I use to access your text)

"Here is the page number: 9705231 Is this enough?".

S20 has received an auto-rejection message to some mail request earlier which the system read as possible auto-processing exchange (many of whom are used by submitters), in this case with a subject line,

"What program should I use to access your text"
...but missing any auto-identifiable information.

17:48 **AA**: ***to* S20** (*copy* ***to admins*** (*subject*: RE: (REJECTED) What program should I use to access your text)

"9705231 but which archive? perhaps you want http://archive.domain/help/view"

17:52 **S20**: ***to admins*** (*subject*: RE: (REJECTED) What program should I use to access your text)

"This archive: http://archive.domain/form/**archive**/"

17:57 **AA**: ***to* S20** (*copy* ***to admins*** (*subject*: RE: (REJECTED) What program should I use to access your text)

"okay, **archive**/9705231 but still, what question?"

18:09 **S20**: ***to admins*** (*subject*: RE: (REJECTED) What program should I use to access your text)

"Sorry, in order to access text, such as 9705231.text.ps.gz, what program should I use? (In other words, if it were a Word document the suffix would be ".doc", if it were WordPerfect it would probably be ".wp", if it were Adobe Acrobat it would be ".pdf", etc. However, it does not appear to be any of these. How do I access the text?"

[ignored]