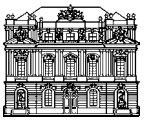
Michael Nentwich

cyberscience

Research in the Age of the Internet

Chapter 6

CYBERSCIENCE AND KNOWLEDGE REPRESENTATION



Austrian Academy of Sciences Press
Vienna 2003

Submitted to the Austrian Academy of Sciences on 10 April 2003
by Gunther Tichy, member of the Academy

British Library Cataloguing in Publication data.
A Catalogue record of this book is available from the British Library.

ISBN 3-7001-3188-7

Austrian Academy of Sciences Press
Tel. +43-1-5129050-3405, Fax +43-1-51581-3400,
Postgasse 7, A-1010 Vienna
Email: verlag@oeaw.ac.at
http://hw.oeaw.ac.at/cyberscience

Layout, cover & type-setting: Manuela Kaitna, A-1080 Vienna
Printed and bound in Austria by Manz Crossmedia GmbH & Co KG, A-1051 Vienna

DETAILED LIST OF CONTENTS

> "We are not dealing with journal articles any more – we are dealing with research communications among scientists. Thus the terms 'article', 'paper' and 'publication' should die. (...) In the new paradigm, research communications are n-dimensional as one can 'jump' from one idea to others."
> (LaPorte et al. 1995)

6 CYBERSCIENCE AND KNOWLEDGE REPRESENTATION

Already in 1945, Vannevar Bush, in his often-quoted article "As We May Think" (Bush 1945), discovered that "professionally, our methods of transmitting and reviewing the results of research are generations old and by now are totally inadequate for their purposes." The age of cyberscience, envisaged by Bush, enables new methods. While there is a long tradition to communicate scholarly knowledge in the form of a (linear) text, i.e. in an article or a book, the increasing digitalisation of texts may lead to a different form of knowledge representation. For instance, the digital medium allows for the inclusion of formats other than text and simple graphics. Scientific publications may soon include, or even be entirely, audio and video sequences. Furthermore, the hypertext technology may enable scholars to present their findings in a modular way and thus provide the readers with multiple ways to access their texts. In addition, in a fully electronic and networked environment it would be possible to link all units of academic output (texts, databases etc.) to form a combined knowledge base, thus revolutionising scholarly products.

This chapter is the first of three chapters dealing with the future of academic publishing in the wider sense. While chapter 7 analyses E-publishing in general, that is the future of print publications, the economics of E-publishing and the prospects of the publishing industry, and chapter 8 addresses the question of quality control in the digital age, the present chapter (6) focuses on the new opportunities for academic knowledge representation and analyses possible scenarios. I shall proceed in the following steps: By way of introduction, I shall look at knowledge representation in academia today (6.1). Then I shall present the cyberscience developments in detail, that is multimedia, hypertext, virtual reality and other new forms of visualisation (6.2). Then, as a thought experiment, I shall develop various scenarios of how knowledge representation might look like in the future. I hypothesise that academics of all types will move from text to databases (6.2.2.1). In a next step, I shall discuss the consequences of this novel kind of knowledge representation. For instance, the inherent dynamic of electronic texts, the changing notion of authorship in a cyber-publishing environment and the impact on the relationship between research and the public will be discussed (6.4). In the final section, I shall analyse the factors influencing the development with a view to assessing the scenarios (6.5).

6.1 Knowledge representation today

Whatever researchers find out about their subject fields and no matter how they arrive at their results, at the end of the day, they are bound to nail it down in some form in order to communicate it to their peers for further discussion. In general, oral transmission is not enough and even conference presentations are eventually written down and distributed. Today, academic knowledge is represented in the following two main formats:

(1) *Texts in various forms*: The by far most common format is the written word in the form of texts. They vary in length: there are very short (that is perhaps only a few paragraphs long) research communications or research notes, articles (or "papers") of several pages and book-length texts. All forms exist both in the traditional printed-on-paper version and in digital format (cf. chapter 7). In general they are fixed or non-dynamic, that is, once published, they do not change anymore. The only exception (in special circumstances) is so-called loose-leaf publications, which can be continually updated by printing new pages to be added or to replace existing pages.

Except for the shortest formats, texts are normally structured by section headings of different levels. How many levels and headings you find in a text depends not only on length but also on discipline. While lawyers, for instance, write deeply structured texts with up to seven or eight levels of headings, philosophers often use only one level and do not even give the heading a title, that is, titles are just consecutive numbers or even three little asterisks. However, "linear texts are generally much less structured than section headings suggest" (Kircz/Roosendaal 1996, 9). By "linear" I understand the classical form of text which has a beginning and an end and which is normally meant to be read in a linear manner, that is, from the beginning to the end. This is not to say that such texts could not be read the other way round, that is beginning with the last paragraph or section, or by jumping from one section to the next regardless of the pre-set path. However, the reader who does not follow the original order has to be careful when complaining about, for instance, a concept or terminology not explained in the section that s/he is just reading because it might have been defined in a previous part that s/he has not yet read.

However, as just mentioned, academics do not always want to *read* a text from A to Z, but they often (perhaps in most cases?) *use* the text in different ways. The most common cause probably is to search for a particular argument or a bit of information. Hence, even traditional linear texts have what one may call hyper-elements, i.e. parts of the text that partially break with the linear order. For instance, academic texts include the following hyper-elements:

- *footnotes* or *endnotes* (which provide for further information and comments on a second layer so that you normally can read the text disregarding the text in these notes);
- *contents lists* and *lists of figures or tables* (which allows you to quickly find particular sections of the text);
- *indices* (e.g. for keywords or cited judgements and their appearance in the text);
- bibliographies (to which the main text refers for the details of the references to other literature and which even refer to other texts outside the present text);
- *marginal text* (which provide the reader with a shortest version of the text beside the main body of the text); and
- *glossaries* (which give definitions and explanations of a set of key terms used in the main text).

However, as the medium is unalterable itself, these hyper-elements are static or "clumsy" (Treloar 1996, 136), that is it is not completely effortless to follow them (you have to often skim through the pages back and forth).

Academic texts often contain tables, figures (graphics) and photographs. As they need to be printed and as colour printing is still very expensive, the latter are, in general, black and white only. At present, however, all three share that they cannot be "dynamic" in the sense that their contents are fixed once printed.

(2) *Databases*: By their specific nature, not all knowledge produced in academia lends itself to be presented in a written text. In particular, data like numerical series, either stemming from experimental measurements or observation, or specifically computed (like prime numbers) are best presented in well-ordered lists and tables. Indeed, there are many thick volumes full of numerals. Other types of information that cannot be represented in a narrative text format are put on huge amounts of cards which, in turn, are stored in files which can fill whole rooms. In particular, bibliographical information has been traditionally stored like this, but also addresses of the persons included in a survey. An important disadvantage of these traditional "analogue" databases is that you have to duplicate the whole data set in case you want to sort it in a different order. For instance, many bibliographical index card files have one set ordered by authors' name and another sorted by title or subject. This has not only obvious limitations as to the various ways possible to order the set, but is also very cumbersome when it comes to updating the database (as you need to do the same updating at many places in the base). Furthermore, searching in analogue databases is as cumbersome and inefficient as in printed text in general.

Since the advent of the computer, digital databases of all sorts have almost entirely replaced the old lists, tables and file cards. The computer allows combining and sorting the same data in many different ways and searching in the full text of all database entries is very convenient. Digital databases can be very simple lists, which are, however, when compared to their analogue predecessors, much more efficient because of the power of digital searching. They can also be sophisticatedly designed multi-layered and complex systems.[570]

6.2 Cyberscience developments

Cyberscience has the potential to go well beyond the status quo of academic knowledge representation as described in the previous section. First of all, the move from the printed to the digital representation needs to be acknowledged (6.2.1). In addition, two further elements are "in the pipeline": multimedia and new forms of visualisation (6.2.2) and hypertext (6.2.3).

[570] See fn. 196 in 2.3.4.

6.2.1 Digitisation

It is not the newest phenomenon, but nevertheless at the very root of what will be academic knowledge representation in the age of cyberscience: there will be only very, very few researchers left who do not produce digital text. Not so long ago, text was produced exclusively by hand and with the help of typewriters. Computers with word processors entered the academic scene only in the 1980s. Still in the early 1990s, most publishers composed manuscripts anew and were not able to take over digital files from word processors for the typographic process. Today, except for very first drafts and notes, even the early versions of most academic publication see the light of day in digital form. For sure, they are printed out on paper several times on their way from first draft to final publication as a book or journal article, but there is always a digital representation of it along the way. Increasingly, even the final state of the publication is the digital form (eventually printed only by the reader, i.e. *after* publication).

This is not to argue that the "digiscripts", as opposed to the traditional manuscripts or typescripts, are of a profoundly different nature. For sure, it is still the same human scholar who conceptualises and drafts the text, and who types or dictates it. However, we should not underrate the impact of word processing on writing. As texts are created more and more "on screen" with only marginal handwritten drafts and less and less frequently by dictation, word processing software comes into play and may influence *how* the text is written and structured (see below 6.4.1.1).

6.2.2 Multimedia – hypermedia – new forms of visualisation

Initially not related to the Internet but to the potential of the computer as such is multimedia. This is "access to text, sound, images, video in an integrated presentation" (Neal 1997, 7), in other words the inclusion of non-text in text or the use of other media independent from text as a novel form of knowledge representation. Today, the Internet begins to play a role here, too, as it is possible to incorporate multimedia material and data from remote sites. This has the potential of revolutionising the way we think about non-textual knowledge representation. We can distinguish the following types:

- *photographs or scans* in multiple sizes and qualities for different purposes, e.g. photographs of a study object (historical document, tablet with inscription) under different light conditions or taken with special cameras;
- *digital audio files* of, for instance, key passages of an interview, of music examples, of historical speeches, of sample text spoken in a foreign language (to demonstrate pronunciation);
- *video clips* of, for instance, an experiment, of an interview situation, of a typical behaviour (e.g. in ethnography or ethology);
- *animated graphics* instead of a series of static graphics to show a development over time or under varying ancillary conditions (e.g. in chemistry or mathematics); and
- *trial software* plus trial data (or even a link to the current data) which can be run remotely to test a claim made by the authors (e.g. in economics).

As all of these elements would, in a digital publishing environment, be interconnected by hyperlinks (see below 6.2.3.2), the combination of the above is often called "*hypermedia*".

This is not to claim that future academic output will only consist of a combination of those new elements. This will certainly vary from discipline to discipline, from field to field. What seems most likely at this stage is that text will stay at the centre in most fields but will be considerably enriched and augmented. However, some argue that text may, in some circumstances, only have an auxiliary function. A substantial decline in the preference for graphic communication as opposed to oral (audio-visual) communication in academia has been predicted (Atkinson 2000, 65). As text and pictures can be mixed in electronic form, we may be reaching a situation where pictures could become primary information sources again, with the text added only to explain the pictures in the model or theory of the author (Kircz 1998a, 6). Text may even play "a less prominent role" (Kircz 2001, 7) and that although "language will remain the essential transfer mechanism for knowledge exchange, non-linguistic communication will regain some of the prominence" (ibid.). Although Kircz is a physicist, the potential for integrating the visual and the textual within the same medium has been acknowledged and also welcomed in other fields, e.g. in ethnography: "It allows us to make the step from thinking of the visual merely as illustrative of argumentation spelled out through the printed word, and to see it as itself constitutive of meaning." (Dicks/Mason 1998, 3.7)

In some respects, the digital surrogate of an artefact may even be better than the original, both for research purposes and for presentation. By digitising you can, for instance, enhance particular characteristics, which you would normally not be able to see (e.g. in astronomy or with regard to stone inscriptions). In addition, digital archives gain through the combination of individual items – something you could hardly achieve in reality, let alone in a publication. In particular, the combination of the digital surrogates in a search space allows for powerful queries which gives the digital item a distinct advantage over the original (Mueller 2000a, 9).

In a "research communication" (LaPorte et al. 1995), the academic "author" is in a position to let the "reader" experience his/her own reasoning. S/he can make the reader see or hear differences or characteristics instead of only describing them.

Furthermore, completely new ways of communicating knowledge and experiencing experimental data, such as *virtual reality* become possible in the new academic working environment. Present day computing already allows for so-called real audio and video via Internet and the simulation of haptic experiences in virtual reality test environments. The author may be "able to simulate the various analog types of perceptions in electronic (binary) form, the reader of the transmitted message can compare his/her own experience of the same sensory experiences, with the interpretation of the originator" (Kircz 1998a, 5). Although much of virtual reality still resides nowadays in the world of games (and estate agents), there is certainly a real potential in it for science as well. Think of an archaeologist who lets his/her "readers" experience the findings of a study with a view to the look and feel of an ancient city. Or consider an ecologist letting us look into the future of an ecosystem based on a scientific model of its evolution.

In addition, there have been a number of further new developments in the area of *visualisation* of results (e.g. OECD 1998, 205ff.). One interesting and promising way of presenting all kinds of (social science) data has been developed by Krempel whose coloured graphical representations of networks shed some light on the prospective future of the visualisation of statistical, as well as of qualitative data (e.g. Krempel 2002).

6.2.2.1 Examples of academic multimedia implementations

Multimedia is not yet widespread in academia (see already 3.3.3). A few examples, however, will suffice to give an impression of its potential. Those fields working with and interpreting original documents are at the forefront. In history, whole archives are being digitised and the documents are being offered to the remote user in various formats and scales with differing levels of details. For instance, the Codices Electronici Ecclesiae Coloniensis (CEEC) project[571] offers original documents of an old church library in Cologne in four different resolutions. Picture archives are available in many other disciplines, too, e.g. in astronomy. The idea is that the items in these picture and text databases can be directly referenced from E-publications. The Blake Archive[572] offers illuminated books, made accessible with sophisticated Java applets allowing for direct comparisons between different versions of the same text. A special tuning device lets the user set the screen resolution to represent exactly the original size. In the bio-sciences, academic articles are often enhanced with additional graphics and colour pictures. One interviewee for this study noted that in some respects the digital pictures are often better than those printed (in particular if only black and white). In addition, he stated that if the publisher allowed for colour prints to be included, the digital versions are still better because they are zoomable. Animated graphics seem to be used so far only in some sophisticated CD-ROM textbooks in mathematics. Links to software used to compute the data discussed in the publication are not exactly frequent, but may be found.[573] Some authors in economics also provide access to the original data they used for computing.[574] A linear text by this author (Nentwich 1999a) has been turned into a "multimedia feature" (Nentwich/Gagliardi 1999)[575] with sound bits, little video-clips and sophisticated navigation tools (see also below in 6.4.1). The "Ästhetisches System"[576] is a representation of an interdisciplinary analysis called a "cross-over" of aesthetics and sociology. It has a so-called "adaptive hypermedia interface" that presents you with text, pictures, audio and a wobbling navigation tool simultaneously. The latter shows you which node (module) you are currently seeing and what surrounding nodes there are.[577]

[571] <Cyberlink=566>.
[572] <Cyberlink=395>.
[573] E.g. in a paper published in the EIoP (<Cyberlink=699>), see fn. 1 of paper 1997-022 by König/Bräuninger.
[574] E.g. Tsebelis (<Cyberlink=775>).
[575] <Cyberlink=776>.
[576] <Cyberlink=115>.
[577] A similar application is "digitalCristal" <Cyberlink=911>.

6.2.3 Hypertext

> "In considering the future of scholarly information exchange, we must therefore take into account (...) the effects of computers on scholarly reading and writing. Certainly one of the best approaches to such an assessment is to focus on the phenomenon of hypertext because it is through the concept (if not yet the reality) of hypertext that we begin to sense the most fundamental and far-reaching effects of the computer on communication in general and scholarly information exchange in particular."
> (Atkinson 1993, 202)

As we have seen in 6.1, hyper-elements in traditional texts belong to the standard repertoire of academic writing. In addition, the tools we use to write them, the word processors or text editors already include the digital relatives of these static elements. The hypertext features of MICROSOFT's WORD 97 are, for instance, the footnotes that pop up on the screen using hyperlinks, the 'document map view' allowing the user to see the structure of the document at a glance and the integrated hypertext functionality that enables the author to set up links to the WWW within his/her text document. Also the so-called "Object Linking and Embedding" (OLE) technology is already allowing for hypermedia effects, as it allows programmes to share data so that, for example, a graphic created in one programme can be accessed through another one (Dicks/Mason 1998, 7.4).

The same feature has recently been put into practice also in an E-book where clicking on the footnote opens a pop-up window with the content of the footnote.[578] This is, however, only the beginning. The potential for making texts more complex and less contained is clearly already with us. As we can all infer from our experiences with the WWW (cf. 2.1.2), the hypertext technology is very powerful and goes, as we shall see, beyond pure "linkability" (Neal 1997, 7), i.e. hypertext connections can be established among diverse and remote information sources.[579]

I shall discuss the various elements of hypertexts below in the following subsections. A hypertext is "non-sequential writing" (Nelson 1992 (1981)) or a non-linear text, i.e. consisting of a number of (text) 'modules' (6.2.3.1) which are connected to each other by so-called 'links' (6.2.3.2). Clicking[580] on links leads the reader to the linked module. I speak of non-linearity because – contrasting standard texts which I call 'linear' – hypertext modules may be linked not only in one sequential form but may, like a network, be interlinked in multiple ways. Accordingly, such a 'web' of hypertext modules can be read in multiple ways or according to different 'paths' (6.2.3.3) since the reader is free to follow the links in the order that s/he prefers.

The idea of large hypertext structures was already born by Bush (1945) although he was not explicitly talking about "hypertext" and called his imaginary machine, which essentially linked textual and non-textual material together, "MEMEX" (from 'memory' and 'text'). He envisaged a "device in which an individual stores his books, records, and communications, and which is mechanised so that it may be consulted with exceeding speed and flexibility". The user would build trails (i.e. paths) in an associative manner. While

[578] <Cyberlink=794>
[579] One of the best introductions to hypertext is still Kuhlen (1991).
[580] While a "mouse" is a very convenient device to access a hypertext, it is, however, not essential. Most hypertext browsers allow the reader-user to "jump" from hyperlink to hyperlink by pressing the "Tabulator" key and to activate the link by pressing the "Enter" key.

acknowledging that indexing certainly has its merits, Bush argued that we are in deep need of more human-like ways of structuring our material. In the 1960s, Ted Nelson coined the notion of "hypertext" and later programmed the famous "XANADU" software, extending Bush's ideas from the individual level to a world-wide hyperbase (Nelson 1992 (1981)). Later, sociologist Luhmann (and many others) put Bush's idea into practice but without any mechanisation (see his description of his famous card file in Luhmann 1992). Krajewski took up this idea with modern means of hypertext technology and built SYNAPSEN[581], a hypertextual card file (Krajewski 1997). Another early milestone was Apple's HYPERCARD system (1987), which gave the user of desktop computers a hypertext editor for the first time. Finally, with the invention of the WWW and the first WWW browser, MOSAIC (1993), development accelerated. Today, through the WWW, hypertext is ubiquitous and known to many outside the hypertext community. There are also authors of hypertext fiction and those analysing them (e.g. Kolb 1994). This chapter will deal with the research related perspectives of hypertext.

In a first step, I shall discuss the three main elements of hypertexts: modules (6.2.3.1), links (6.2.3.2) and paths (6.2.3.3).

6.2.3.1 Modularity

While putting hyperlinks into a text makes it less "linear", it does not necessarily alter its fundamental structure. The text may well stay a linear, however "enhanced", perhaps more "user-friendly" text. The above-mentioned hyper-elements in traditional print text do not alter the fundamental character either, even if turned into electronic hyperlinks. In principle, however, the hypertext technology also enables the author to write something different: the non-linear or modularised text. It consists of a large number of "modules" (or "nodes"), linked together with hyperlinks. St. Laurent (1992) speaks of "creative fragmentation" to capture this concept of modularity.

Hence, modules are the basic units of hypertexts. A module or "lexias", to borrow a term from literary criticism,[582] is a piece of text with one or more links leading to it and from it to other modules. Modules may be longer or shorter, depending on the type of hypertext. We can distinguish between the following two basic types: fixed modules (1) and free modules (2).

(1) *Fixed modules*: In a highly standardised publishing environment it may be possible to distinguish between a number of rather technically defined parts of a scientific text. These may include, for instance, the abstract; the description of the state of the art or status quo of research so far; the experimental outline; the description of the technical details of the machinery; the results; the analysis; the bibliography etc. These may be the fixed modules of any hypertext in that discipline (similarly Atkinson 1993, 208).

The advantage of modularising the texts this way lies with the possibility to link the corresponding parts of different articles with each other to form new ways of presenting the research. This can be done in two basic forms: either authors do not write those parts of the paper that have already been explained elsewhere again, e.g. the experimental outline, the technical description or the status quo. Or they create new forms of articles focussing on specific aspects of a field, e.g. the methodology only or the development of the status quo by linking several such modules together, perhaps with additional commentary.

[581] <Cyberlink=69>.
[582] Others speak of "fragments" or "nodes" of text (e.g. Davenport/Cronin 1990, 176).

With a view to academic publishing in experimental physics, a research group at the University of Amsterdam developed such a scheme of modules (and links) (Harmsze 2000). The Amsterdam group distinguishes between elementary and complex modules. Based on a typology of scientific information in their field, they characterise the information in four ways contained in the modules (Harmsze et al. 1999):

INFORMATION CONTAINED IN SCIENTIFIC ARTICLES

- Domain-oriented (keywords)
- Bibliographic data
- Range of information
  - Microscopic (information that belongs only to one particular article)
  - Mesoscopic (functions at the level of an entire research project)
  - Macroscopic (transcends the level of the research project)
- Conceptual function
  - Positioning (situation; central problem)
  - Methods (experimental; theoretical)
  - Results (raw data; treated results)
  - Interpretation (qualitative; quantitative)
  - Outcome (findings; leads to further research)
  - Meta-information (bibliographic, abstract, references, acknowledgements, map of contents)

Overview 6-1: Modular structure for electronic scientific articles in physics according to Harmsze et al. 1999

The above structure can be used to typify modules that have to be written in the course of composing a scientific article. Indeed, Harmsze (2000) proposes such a modular structure and exemplifies it with modularised versions of two sample articles from experimental physics.[583]

(2) *Free modules*: In contrast to the model of fixed modules, in the alternative the text is not only modularised on a much smaller scale (e.g. paragraphs, thoughts, arguments), but also regardless of the module's structural position at the meta-level. Free modules can be thought of as the units of a very fine-grained text structure, that is, of a very "deep" series of headings (which you would normally not do in a linear text[584]). At the extreme, each paragraph or sentence represents an individual module. The author has to decide what the "knowledge unit" is which has to go in a separate unit. Perhaps, s/he will be guided by the rather pragmatic principle of what should be individually addressable. Only those thoughts and arguments contained in a separate unit may be "re-used" in another context (or referenced to). Free modules may also be combined with fixed modules.[585]

[583] Online accessible, see <Cyberlink=774>. An alternative to Harmsze's model would be a hierarchy of modules, e.g. first raw data, then data-reduction and handling, then data-representation (Kircz/ Roosendaal 1996, 10).

[584] Perhaps with the exception of legal scholars commenting a statue. Another such example is Wittgenstein's "Tractatus Logico Philosophicus".

[585] The above mentioned text by this author (Nentwich 2000b) <Cyberlink=31>, for instance, is a hypertext example using mainly free modules, but also a number of types of modules which could be fixed (such as acknowledgements, abstract, bibliography, contents list etc.).

6.2.3.2 Links

Links are text-markers – i.e. highlighted, underlined or otherwise marked text or small pictures such as "icons" or "buttons" – which the reader/user can click on with the PC mouse (in case s/he uses a graphical user interface) or by selecting it with the keyboard (in a pure text environment). As soon as a link is selected (i.e. pressed or clicked on) the hypertext client "jumps" to the target indicated in the description of the link. There are

- *internal* links to other parts of the same document (hypertext) – this is the electronic version of traditional hyper-elements; and
- *external* links, e.g. to quoted literature, documents, data, software etc. ("resource links" in the words of Armstrong/Lonsdale 1998, 34) residing outside of the main text, that is in the wider hypertext environment – this is something not possible in printed text. Furthermore, there may also be links to "companion web sites" (ibid., 36) with more resources, references ("further reading") etc., either by the author or by the publisher, a link to an author biography, also bookshop links. A novel feature of E-journals, not possible in the print medium, but only in a dynamic hypertext environment, are *forward links*. They come in two forms, either as "inverse citations" (Raney 1998, 7), i.e. links to subsequent published articles citing the current article[586], or as "forward pointers" which update articles with citations to other material published later (Fisher 1997, 4).

Technically speaking, a hypertext may only use external links since all modules may be separate electronic files. In terms of content, however, most links may nevertheless be internal since all modules form part of the same "virtual" document. (In some scenarios, however, even the above technical distinction will become blurred.)

From a syntactic point of view, we may further distinguish between:

- *explicit* links (or citations), i.e. those readily available to the reader, suggested by the author, and
- *implicit* links (or citations) the reader can make "by looking for similar texts that use the same signifiers (words, sounds) in similar sequences or proximities in other files accessible throughout the network" (Atkinson 1993, 203). Using the hypertext somehow creates these links.

Furthermore, links may be *bi-directional* or *unidirectional*, that is the module at the other end may or may not have a link back to the original module (Davenport/Cronin 1990, 177). Most internal links will be bi-directional whereas external links can only be bi-directional under very special circumstances as it involves a change of the target module – which involves some control over extraneous material.[587]

In a standard web document full of links (e.g. lots of blue underlined words) the reader is not in a position, from the appearance of the link, to know what to expect when clicking on such a link. These standard links carry no "meaning". More sophisticated hypertexts will, however, have meaningful links.

[586] In the large bibliographic database of ISI, Web of Science (<Cyberlink=488>), this is already implemented.

[587] See, however, the activities in the framework of the Semantic Web initiative to allow for shared and individual annotations to web documents (see 2.3.2). A related issue is that links are an important topic within the XML activities under the label of XLL (see 2.1.2).

Meaningful links

Links may become a proper information object with clear characteristics telling the reader what type of information is to be expected when following this particular link. Most web sites already use some typified icons for *navigation* on the website. These navigation links are the most primitive meaningful links. For instance, a small arrow oriented towards the right side tells the reader that this is a sequential path leading to the next module in a prefixed row; a little house normally means that this is the direct way back to the start page; and an icon with a question mark or a magnifying glass means "This is the way to the in-built search machine of this site." Icons may thus typify certain relations between the modules linked together. This can be much expanded to help guide the reader/user through the content of a hypertext.[588]

Going beyond navigational links, a special type of meaningful link is one leading to a *comment*. Unlike non-electronic texts, hypertexts (and to some degree also linear E-texts, e.g. in some E-journals) can include comments. A comment may be a counter or supporting argument, an example, a hint to further literature, to another hypertext module etc. Comments would be particularly useful in my scenario 3 (see below 6.3.3) in which I explore the possibility of 'hyperdiscussions'.

There is, however, potentially even more than only navigating and commenting. Links may also tell the reader about the content of the module to be reached. The relations between modules are in itself information objects (Kircz 2001, 9). For instance, we may classify links according to rhetorical grounds (dissension, corroboration, explication, see Davenport/Cronin 1990, 177). The Amsterdam physicist team mentioned above put much effort into establishing a list of meaningful links. They distinguish between "organisational relations" and "scientific discourse relations" (e.g. Harmsze et al. 1999). The formers "express the organisational coherence of the modular network", the latters, by contrast, would allow authors to indicate why they refer to another module (for details see Overview 6-2 below).

ORGANISATIONAL RELATIONS	
Hierarchical	Asymmetric relation between complex modules and their constituent modules
Proximity-based	Symmetric relation between linked modules expressing whether they are part of the same collection
Range-based	Asymmetric relation expressing the difference in range between linked modules
Administrative	Asymmetric relation between conceptual modules and the module representing their meta-information
Sequential	Asymmetric relation between modules linked to form a complete or a more easy-going reading path
Representational	Asymmetric relation between different representations of the same information

[588] The only typified links implemented in Nentwich (2000b) are those in the 'Paths' section (back/forward; contents; home; short path).

SCIENTIFIC DISCOURSE RELATIONS	
Communicative function	Targeted at increasing the reader's understanding of the message or at increasing his/her acceptance; they can be further divided into elucidation (increase the reader's understanding) and argumentation (increase the reader's acceptance of a standpoint)
Content-related	Targeting at making clear the relationship between the content of arguments and other pieces of information; they can be further divided into those signifying a relationship of dependency in the problem-solving process, elaboration, similarity, synthesis or causality

Overview 6-2: Meaningful links according to Harmsze et al. 1999

Given this great variety of relations (I have not listed all sub-categories to be found in Harmsze's writings), it seems difficult to implement this in a simple way, e.g. with icons, since obviously you would need both small pictograms (which cannot carry much information) and quite many different ones. Avoiding confusion on the reader's side is certainly a difficult task.

Even a "discourse-mark-up-language (DML)" has been proposed in sociology, in particular with a view to making discussions in the net more efficient. Rost (1996b) based his idea on the observation that participants in E-mail discussions already use various "tags", i.e. markers, to indicate a certain meaning of the pure text: 'smilies' or 'emoticons' (e.g. :-)) and, sometimes, even pseudo-tags such as

<sarcasm on> a provocative statement .... <sarcasm off>

Similar to HTML which gives the Internet browser instructions about how to layout a particular text, he proposes the development of a set of tags which would give the reader and also the "artificial reader", i.e. knowbots or intelligent agents, hints as to the meaning of the content. Based on linguistic categories, Rost distinguishes between three different kinds of contributions to a discourse (see Overview 6-3) and proposes DML tags with three parts. The first component would signify the relationship between the statement and other statements, the second the relationship between the statement and the facts in the world, and the third would tell us something about the assessment of the author. This schema would lend itself to being transposed into specific links between the various contributions to an academic discourse.

CONTRIBUTIONS TO ACADEMIC DISCOURSE:	
<Präsuppostionen>	Would give the context of the following text, e.g. that it belongs to a certain (sub-)discipline
<Illokutionen>	Would relate to communication and mark the links: approval/disapproval; encouragement/discouragement; doubt; recommendation; question/answer etc.
<Propositionen>	Declaration; statement; observation; reminder; definition, comment, explanation, correction; follow-up; abstract etc.

Overview 6-3: Types of discourse tags according to Rost 1996

Meaningful links can be constructed in different manners. Another proposal (St. Laurent 1992) suggests the distinction between

- the *active* link model where the editor must choose each link and create starting and ending points for each connection (that is "by hand"); and
- the *passive* link model where the editor or author has to assign each module one or more keywords; the software automatically creates links between those modules with the same keyword. "Keywords offer the creator a chance to let the text link itself" (ibid, 4).

For sure, in both the active and the passive version, "much of the usefulness of hypertext is dependent upon the skills of its creators" (ibid, 4).[589]

6.2.3.3 Paths

Modules are combined by links providing for reading paths. The simplest form of a hypertext resembles a "string of pearls", i.e. a sequence of modules with one link in each module leading to the next module in the row. Obviously, this does not explore the full potential of hypertexting. As soon as the modules include more than only one link leading to a next module, the various modules of a hypertext can be read in different orders.

The author may propose paths to the readers, e.g. by indicating a sequential path. If the user has finished reading one module s/he simply clicks on "FORWARD" (or the equivalent icon) and will be led to the next module and eventually through a big proportion of the hypertext (but not necessarily the whole hypertext, that is, through all modules, as the author may choose to open "cul-de-sacs" or dead ends which are not touched by the main route). Here, the author somehow mimicks the structure of linear texts. There may be paths for experts in the specific sub-field or for scholars from the same (wider) discipline, paths for science journalists, for lay people, teachers, students etc. (e.g. Davenport/Cronin 1990, 185). Each of these paths will lead the reader to a different set of modules of the hypertext and let him/her explore it in a different order. One of the authors of the E-journal "Living Reviews of Relativity"[590] has actually used the fact that articles are broken down into different sections, each being a separate HTML document, to create a document that could be read in different ways by different users (Wheary et al. 1998, 5).

Most importantly, the reader is not bound by the proposed paths and can explore the hypertext as s/he pleases. This is exactly one of the major strengths of hypertexts, in particular of large hypertext structures as discussed below (6.3): the larger a hyperbase is, that is the more modules there are, the greater the possibilities to combine modules. For sure, orientation will be much easier when systematic use of meaningful links is made (see above).

6.2.3.4 Examples of academic hypertexts

Experiments with and implementations of hypertext are still even more rare than multimedia applications in academia (see above 6.2.2.1). There are, however, a few examples showing the potential of this for scientific publishing. The two elaborated examples of hypertext articles (with fixed modules) in experimental physics have been mentioned already[591]. In the social sciences, there are hypertexts with free modules by Ingraham

[589] See above 5.1 on the multiplication of author's tasks.
[590] <Cyberlink=237>.
[591] <Cyberlink=774>.

(2000)[592] and by Nentwich (2000b). The "Ballad of the Internet Nutball"[593] is an experimental hypertext Ph.D. thesis in English literature. The "Tree of Life"[594] is a collaborative hypertext database with information about phylogeny and biodiversity. The Valley of Shadow[595] site is an impressive hypertext archive on the history of the American civil war.[596] Wingert (1996) describes an experimental hypertext on the basis of a transcript of a speech given by the philosopher Vilém Flusser.

6.3 From text to knowledge base – one plus five scenarios

Since hypertexts offer such great potentials, this chapter builds on the assumption that hypertext and hypermedia will gradually become the standard ways of representing academic knowledge. Classical linear texts are already on the retreat in some disciplines as hyper-elements become increasingly digitised and dynamic. Already today, the large P+E-journal databases of commercial publishers are more than simple collections of linear texts as there are multiple links between the different items. In particular, links to quoted articles and (dynamic) links to "related articles" (cf. 2.3.4.1) produce a web of academic publications. If hyper-citations or "actionable citations" (Davidson/Douglas 1998, 8), i.e. direct hyperlinks to quoted literature, become widespread, we may see a net of interlinked E-papers emerging. "(A) growing body of scholarly literature may be enmeshed in a net of links and connections that multiply the value of each item appreciably" (Okerson/O'Donnell 1995, Conclusion). Publishers are increasingly co-operating, which allows cross-linking between journals of different publishers. Similarly, the providers of bibliographic databases are trying to offer cross-references in all possible directions. Both the articles and the links are stored in dynamic databases, which are increasingly interconnected (cf. 2.3.4.3). In addition, many further types of links between the various representations of academic knowledge are more and more widespread. With Hunter (1998, 2) we may distinguish between:

> "ways to link journals electronically:
> - from bibliographic searches to the article's full text;
> - from tables of contents services to the full text;
> - from a cited reference to the abstract or the full text of the cited article;
> - from the article to later citations of the article via a database of journals aggregated for searching;
> - from the journals of one publisher to those of competitors; and

> ways to create other intelligent, useful links more specific to the research in the article:
> - connecting gene sequences in articles to detailed records in gene databases;
> - upgrading 2-D chemical formulas to 3-D chemical structures;
> - linking experiment descriptions to the underlying large data sets;
> - expanding from text to audio, video, animation;
> - moving to related Web sites, information, news, chat, ads."

[592] <Cyberlink=118>.
[593] <Cyberlink=30>.
[594] <Cyberlink=219>.
[595] <Cyberlink=295>.
[596] For further examples in history and elsewhere, see <Cybercategory=20>.

My *base line scenario* therefore is that, based on the current status quo and dynamics, a worldwide virtual hyper-database of academic publications that are still linear will soon be realised. The database will be virtual, as the user retrieves all publications through different interfaces and the publications are stored in different locations, but the effect is the same as if they were all accessible in one place. At the end of the day, if the researcher's institution has bought licences for all relevant journal databases, s/he would be able to access all articles through links leading from the reference list to the full text, and from the latter's reference list further on. This has been called the "immediacy" of this new referencing capability (Dewar 1998, 18). In E-journal publishing "the application of citation linking presumes from the outset that ultimately every citation will be linked to the cited source in some form and builds on conventions long established in print" (Hitchcock et al. 1997c, 14). An advanced linking strategy would go even further as it would define "relevance as all representations of the same document, all documents that refer to them and any documents they refer to" (ibid.) Once established this feature may be so powerful that it will be almost mandatory and will transform journal usage (ibid.). The only limitation likely to stay forever is that many publications will never be available online with the consequence that not all items in a reference list would be linked. This concerns traditional books as well as older journal articles that might for economic reasons never be made available electronically.[597] In addition, external links will become widespread in E-publications.

A helpful distinction should be made between *"strong" and "weak" hypertexts* (Wingert 1996, 113). The formers use the idea of a network as the central organising principle for the content of the hypertext. The latters only show some of the characteristics of linking. The development towards worldwide cross-linked academic journal databases as just described in the base line scenario is a typical example of a weak hypertext structure. There is no doubt that most academic publications will soon be inter-linked in such a weak (but nevertheless very useful!) manner. However, the potential of hypertext goes well beyond this. My hypotheses about the future of academic texts (to be discussed in the following) include hypertext and hypermedia in a strong sense. There is the chance that articles will not only be linked together, but that they will be hypertexts themselves. I am not talking here of internal linking between the contents list of an E-journal article and its various headings and footnotes, but of modularising articles in a much more sophisticated way (as explained above).

Only on the basis of free and/or fixed modules, may a text become a genuine hypertext with different paths and the potential to "grow together" with other hypertexts to form a web of knowledge units and, eventually, a hyper(data)base.

A few authors have already speculated about this perspective. In the age of innovative E-journals, documents may evolve into databases if they include different paths, graphics and movies, computer software etc. (Wheary et al. 1998, 5). Indeed, the Living Reviews offer an innovative reference database including all referenced articles for all papers published in that journal. As word-processing powers expand, the single-authored text will probably become "more 'leaky' at the edges" (Dicks/Mason 1998, 7.4) that is connected with other texts and resources. One possible outcome of the current growing to-

[597] Note, however, that there are a number of initiatives with the aim to retro-digitise both older books and journals, e.g. History E-Book (<Cyberlink=246>) and JSTOR (<Cyberlink=322>). It seems unlikely that these and further initiatives will ever digitise the whole corpus. See however the proposal to at least retro-digitise everything quoted in 11.2.2.3.

gether of academic publications may be what Dewar (1998, 18) calls the "Internet book". It seems that over the last decades the elemental information product has become more granular. Therefore, the E-encyclopaedia is particularly attractive because it allows one to access directly a short, focused package of information with links to more. Consequently, authors may "be drawn to package their products in small bundles embedded in a large database with links to other elements of the database with related information" (Getz 1997, 11). This would probably lead to still more granular information and the database may become the dominant unit of trade in academic information.

In the following sub-sections, I shall push the idea even further and lay out a number of scenarios about the inherent dynamics of strong hypertexts.[598] I shall discuss the cooperative dimension of hypertext authoring as well as the potential, that hypertexts of the same author(s) or different (groups of) authors may grow together. This may eventually lead to a situation in which the boundaries between the (hyper)texts vanish altogether. In section 6.4, the likely consequences, in section 6.5 the likelihood of these scenarios will be discussed.

HOW HYPERTEXT MAY AFFECT ACADEMIC PUBLISHING	
Base line scenario:	Web of linear articles
Scenario 1:	Layered E-publications
Scenario 2:	Field-wide thematic hyperbases
Scenario 3:	Open disciplinary hyperdiscussions
Scenario 4:	Hyperbooks
Scenario 5:	Consolidated knowledge bases

Overview 6-4: Hypertext scenarios

6.3.1 Scenario 1: "Layered" E-publications

The modularised structure of hypertexts together with multiple paths provides an opportunity to have several *layers* in an article, each consisting of different parts. The entire E-publication is more than each individual reader would see at any one time. There might be both different variants of the same module (in different languages, technical or non-technical) and different paths leading the reader to a different sub-set of all modules included in this particular hypertext article.

This basic idea can be implemented in different manners. One can speak of 'concentric stratification', i.e. a form of presentation "not in the traditional linear sequence, but rather as a set of linked or self-citing levels or strata" (Atkinson 1993, 208). Each succeeding level would include the information contained in the previous level while in addition providing greater degrees of substance and detail. This would establish a new sort of hierarchical structure of formal academic publishing, one that goes beyond the current distinction between monographs and articles.

[598] Note that I do not use the notion of scenario here in the technical sense as developed in future research on the basis of expert-generated scenario techniques. Rather scenario is deployed here in the sense of a possible development path and a partial vision of the future development.

For an E-book, e.g. in history, a structure of "layers arranged like a pyramid" (Darnton 1999, 9) could look like in Overview 6-5 below.

E-BOOK LAYERS	
Core text	Concise account of the subject, available perhaps in paperback
Elaboration of specific aspects	Expanded versions of different aspects of the argument, not arranged sequentially as in a narrative, but rather as self-contained units that feed into the core text
Documentation	Possibly of different kinds, each set off by interpretative essays
Theoretical or historiographical layer	With selections from previous scholarship and discussions of them
Teaching modules	Pedagogic layer consisting of suggestions for classroom discussion and a model syllabus
Debate	Readers' reports, exchanges between the author and the editor, and letters from readers, who could provide a growing corpus of commentary as the book made its way through different groups of readers

Overview 6-5: History E-book layers according to Darnton 1999

Although not explicitly talking about hypertext or paths, Getz proposes an interesting system with similar features. In this system, authors would upload their papers to an E-pre-print server. Multiple editors would select papers and encourage authors to develop the work for different audiences, for instance for a top academic journal or for a wider audience. The different versions of the work might appear in a common database of articles and link to longer versions of the work, to numerical data sets, bibliographies and other related material. The published essays will then be "front-ends to a deeper literature available on the Net" (Getz 1997, 12).

All three ideas focus on different aspects as regards the varying potential to represent the same material. While Atkinson focuses on the amount of material included at each layer and hence on the length (or depth) of the text at each layer, Darnton and Getz have different readers for each layer in mind.

What they all have in common is that the individual academic article of the future would not be self-contained, but part of a layered structure with different, but connected material and/or different ways of presenting it at each layer. The following scenarios will go beyond the individual academic publication and explore the potential of hypertexts to grow together.

6.3.2 Scenario 2: Field-wide thematic "hyperbases"

We may think of (sub)disciplines whose academic papers have, as a rule, always the same structure, i.e. they include the same parts. In the natural sciences this seems common (see the example of experimental physics described above, 6.2.3.1). Even in the social sciences and partly also in the humanities, we find typical components and similar structures. For example, some editors require the authors to deliver not only abstracts, bibli-

ography and keywords, but also particular one-paragraph sections like 'Issues', 'Relevance' or 'Context'. Often also there are 'Methods' and 'Theory' sections and distinct parts presenting the 'Empirical data' and the 'Analysis'. In the age of cyberscience, E-publications may evolve into hypertexts with fixed modules. The modules would be cross-linked, both internally (that is between the different modules of the present author(s)) and externally (that is with further modules of the same author(s) or of different authors. The result would be large "hyperbases" of many inter-linked modules in a given area. Readers would read in the whole hyperbase rather than in individual papers as the many links between the hypertexts forming the base would easily lead them to connected modules originally not belonging to the article with which they started.

Breaking apart the linear text into independent modules, each with its own unique cognitive character, may be advantageous. Most readers are looking only for parts of the information stored in research articles. An almost natural consequence of the split between storage and presentation[599] would then be the separate storage of unique pieces of information (Kircz 1998b). Consequently, instead of being forced to filter through whole articles on the pursuit of particular bits of information, the reader is in a position to find the latter in a much more targeted way. Also the search-engines, databases and knowbots of the future will be able to address particular sections of a hypertext since they "know" what a particular section is about due to the meta-information which comes with the module. For example, if you search for raw data on electoral behaviour all over the world, this may be only a mouse-click away as soon as the relevant section in all E-papers are marked in a special way. Academic hypertexts whose modules are enhanced by meta-data may eventually form the academic part of the Semantic Web (cf. 2.2.2.1).

Another advantage as compared to the status quo (and also to scenarios 3 and 4, see below) is that if this vision becomes reality, it will be much easier to reference to the then standardised parts of future E-papers. If we take, for instance, the methodology sections, authors may refer to a suitable description of the methodology applied in their paper (adding only what they think is missing there). Hence, redundancies will be diminished and duplication avoided. According to Denning (1995), the ACM Digital Library was intended to become such a hyperbase. Already by then, he had observed an emerging "standard practice to think of papers as collections of objects (sections, paragraphs, figures, tables, pictures, and the like) rather than simply as texts". The ACM Digital Library would allow authors "to embed pointers to, rather than copies of, objects in their works" (ibid., 10).

6.3.3 Scenario 3: Open disciplinary "hyperdiscussions"

While many discussion lists are rather a forum for exchange of information, there are others whose participants are really engaged in debate (cf. 2.4.2 and 3.3.5). Their contributions refer to each other and there are so-called "threads", i.e. clusters of contributions dealing, at least initially, with one particular subject. In most cases, there are moderators who try to (re-)arrange the contributions, sum them up from time to time, separate themes and give new impetus to the flow of arguments and counter-arguments.[600] Pres-

[599] The presentation is only realised in the moment the information is accessed (for instance with so-called style sheets, cf. 2.4.3).

[600] This is what Harnad (1990) calls "scholarly skywriting" (cf. 7.2.4.4).

ent day list-servers are, however, not able to present the contributions (modules) in more than two ways: either hierarchical, i.e. sorted by subject (thread) or by the time the contributions were posted. Both types cannot cope with more complex relationships between thoughts and arguments and the reader is left unguided. The same is, by the way, true for bulletin boards on news-servers as well as for present day Internet groupware.[601]

Hypertext applications may be a good tool to cope with the problems of Internet debates since they may be used to "discuss". As with E-lists, the author and the date and time of the contribution are recorded, but the contributors are in a position to place their comments, additions, amendments, pro or counter arguments, examples or whatever exactly where they refer to with vice-versa links. Readers may follow the whole discussion on individual paths, but also sequentially (i.e. in the order of time). In contrast to pure hierarchical representations, discussants may link their arguments or examples etc. to different other threads (that is groups of modules). Here it would be particularly useful to have a common set of typified links representing the "scientific discourse relations" (Harmsze 2000) or a scientific discourse-mark-up-language[602] (Rost 1996b).

The (dynamic) result of such a text-oriented discourse would be a snapshot of the ongoing discussion on a particular topic. We may call them "hyperdiscussions". Compared with scenario 2, this one is a rather "anarchic" approach. There are no strict rules as to what comes into what module. The author ("poster") is free to divide the modules as s/he pleases. This scenario may nevertheless have some standardised elements, in particular with respect to the links (see above). In contrast to scenario 2, cross-linking between the hypertexts of different authors would not lead to a "hyperbase", i.e. a structured database in hypertext format, since the modules are not standardised.

Such a scenario resembles some of the proposals made in the literature. The historian St. Laurent (1992, 4), for instance, envisages multi-author collaborations of a sort of asynchronous electronic conference taking place in well-designed hypertext environments. The sociologist Rost (1998a) complains about the, in his view, deplorable state of the practice of the academic discourse and thinks that a better structured discourse may only be realised on the basis of modern communication technologies. Apart from evaluative tools like scoring servers[603], he demands improvement of those procedures aiming at stabilising the results of the discourse in "a canonical, accepted, encyclopaedic corpus of reference" (1998a, 4, transl. MN). In this respect Rost proposes the further development of hypertext tools such as HYPER-G[604] in combination with a DML (1996b).

6.3.4 Scenario 4: "Hyperbooks"

While the previous scenario is about open text bases – open in terms of participants and in terms of time – which are produced in a weakly organised mode, we could equally imagine more structured enterprises aiming at producing closed, but nevertheless dynamic academic products.

[601] An interesting new development here is OpenTheory (<Cyberlink=50>) which tries to provide for a collaborative development bed for developing texts the same way as open source is developed.
[602] DML, see above 6.2.3.2.
[603] See Rost (1998c) and 2.4.2.
[604] Now: HYPERWAVE (<Cyberlink=43>).

In general, present-day books are outdated the day they appear on the market. The reason for this is simply that the academic production does not stop on the day the manuscripts are handed over to the editors and publishers. Traditional texts are fixed, non-dynamic. The only (rather cumbersome and second-best) solutions are either to have new editions in shorter intervals or to publish loose-leaf editions instead of bound books and have new pages sent frequently to the subscribers of the book.

Dynamic, electronic hypertexts may again be a solution. It is conceivable that single authors or groups of authors decide to continually work on a (common) 'hyperbook'. New information could be added almost instantaneously, thus making the product always up-to-date. In addition, multimedia elements could be included without much effort. Apart from the fact that the contributors would have to master a new medium, this scenario does involve organisational challenges similar to those known from collaborative book projects in the paper world.

6.3.5 Scenario 5: Consolidated knowledge bases

While the hyperbooks described in scenario 4 are the equivalent of today's books or edited volumes, the author's intention may also go beyond the single book. Similar to traditional encyclopaedias, textbooks and commentaries, groups of authors may work together to establish a knowledge base with a specific thematic focus, aiming to assemble the available knowledge in the area. The overall aim would be to provide for a large hypertext presenting the "standard model" or "state-of-the-art" of a sub-discipline, that is the consolidated knowledge of a speciality, including examples, empirical evidence etc. Like in co-authored volumes the authors would have to find some sort of mechanism in order to solve disagreements as to the content of particular modules among themselves. In the long run, there is even the possibility that all hypertexts in a discipline grow together and eventually form a common hyperbase or "consensual knowledge base" (Euzenat 1995).[605]

In case the knowledge is fairly standardised, like in medical literature reporting clinical trials, new knowledge could be published as both prose discussion (that is as traditional linear texts in article format) and as entries into knowledge bases (which Sim 1998, 10 calls "trial banks"). Artificial intelligence would be needed to make these knowledge bases interoperable and scientific publishing could then be a new demonstration area for knowledge-sharing technologies (ibid., 11). Others expect a more hybrid form of representation which would allow both the "collage principle", i.e. the hypertext/database structure, and "the more authored kinds of meta-narrative to be integrated within the same medium (or rather within the same multi-media form)" (Dicks/Mason 1998, 7.4). While not concerned with hypertext in particular, the "Designated Channel" (Atkinson 2000, 67f.) attributes the specific task to administer and fill these knowledge bases to the cybrarians (cf. already 5.3). In this model, all publications in a discipline go into an electronic (field-wide publication) channel. Normative meta-data based on access data (use tracking)[606] and certification (through peer review) is attached to each item. On the basis of an author-provided abstract, the information providers (cybrarians) produce a "syn-

[605] Euzenat thinks that such knowledge bases may become the nucleus of distributed collaboratories (1998).
[606] Cf. 8.2.2.3.

opsis", that is a description of how the certified publication relates to all of the other publications that have been certified in the channel, i.e. the certified whole. The purpose of the synopsis should be "to stipulate what is new or unique in the publication" (ibid.) From this "a new form of cumulative metadata, [called] the surrogate whole", either by the cybrarians, as Atkinson proposes, or by the scholars. This surrogate whole would provide a summary of the discipline or an overview of all aspects of the discipline, which would permit the reader "to move selectively from the surrogate whole into the certified whole to read particular publications". In Atkinson's model, the traditional bibliographical model would shift to the model of encyclopaedia. "The primary purpose of information services must be not to list publications by name, but rather to provide a narrative context of their content" (ibid.).

In contrast to the scenario 4 model (which still reminds us of traditional book projects), the knowledge bases are projects of a field or discipline, hence they would not be the product of a limited group of authors but an "open" project. Everyone (or at least everyone who qualifies according to some criteria, such as being at least a graduate student, or at least a Ph.D. candidate, or being a member of the respective scholarly association etc.) might contribute to the steadily evolving/growing hypertext (cf. also Rost 1998d, 8).

Quality control would be crucial in such a project: How can it be secured that the modules really represent the state-of-the-art knowledge? Various forms are conceivable: traditional peer review, open peer commentary or rating (cf. chapter 8). If the evaluating discussion takes place in the form of added comments, counter-argument etc. (perhaps as in scenario 3), it will be necessary to establish some sort of decision-making procedure to consolidate and close a discussion. By this token the module(s) debated could be amended and the old modules and units generated during the discussion could be moved to some sort of archive.

6.4 Impact assessment of cyber-knowledge representation

Now that I have reviewed the opportunities for knowledge representation on the horizon and drawn some more or less realistic scenarios[607], I need to assess the possible impact of this development. The following sub-sections will go beyond a simple list of possible benefits or problems. By contrast, I shall systematically review all areas in which the shift from analogue to digital knowledge representation may have consequences. I start with an analysis of the changing character of the academic "text" (6.4.1). The next two sections are devoted to authors (6.4.2) and readers (6.4.3) of such new texts. The concluding section discusses the hypothesis that the academic paper is slowly developing into a research communication (6.4.4).

[607] In 6.5, I shall finally analyse the likelihood of these scenarios.

6.4.1 Changing character and type of texts

My first observation is that cyberscience will change the character of academic texts. This relates both to the production process (6.4.1.1), to the length of a publication (6.4.1.2), to the inherent dynamic or fluidity (6.4.1.3), to the intrinsic logic of (hyper)text structures and the coherence of arguments in a modular environment (6.4.1.4) and to the relationship between representation and its object (6.4.1.5).

6.4.1.1 Digital production

> "The word processor, what ever some of its users may say, is *not* 'just a better typewriter'. It signifies a change in the relationship between the author and the language."
> (Birkerts 1997, 212)[608]

As we have already observed, one of the main characteristics of knowledge representation in the age of cyberscience is the fact that it is all-digital (6.2.1). This has a number of consequences on the product level, which I shall address in many of the following subsections. Here I focus on the parameters of the production process itself.

Academic knowledge representations are now produced by digital means, in particular with the help of word-processing software. These production tools are likely to influence the form and content of the produced article. For sure, it makes a difference whether an academic field is text-oriented (that is mainly producing narratives and arguments), or whether text is only a representation of the main results (e.g. of experimental measurements). However, *how* you present your results is always important. Even if some researchers produce text only at a very late stage of their research, they nevertheless have to do it. In any case, in those fields where text production starts very early in the research process (in most of the humanities and social sciences) or is even almost the only activity (e.g. philosophy, literary studies) and where text does not only serve the purpose of presenting data but has the function to convince the reader exclusively on the basis of arguments, changes to the process of text production are likely to have an impact on the nature of the text itself.

So, what distinguishes the digital text production process from its precursor, the age of paper and pencil? Here are some answers:

Drafting: The paper manuscript could not be changed (corrected, amended) continually without it becoming more and more illegible. Rubber eraser and correction fluid have their obvious limits. Hence, in order to see the present status of the manuscript, the only solution was to transcribe it, either by hand or with the typewriter. This was cumbersome and time-consuming. Today, authors are always in possession of a correct and up-to-date representation of their texts-in-the-making. Printing them out is not even necessary as you can check the text on screen. In any case, the number of drafts and versions of each academic publication has certainly increased. This is true both on the "macro" level, that is, the versions of the whole article or chapter, and on the "micro" level, that is, the individual sentences. It is likely that, in anticipation of the obvious limitations of the old eraser technologies, sentences were written down, in general, in a much more mature stage. As it is now very simple to delete words and groups of words, to rearrange their order, to replace them by synonyms, to exchange paragraphs etc., authors do this more frequently.

[608] Transl. MN; emph. in orig.

Structure: Word-processors are able to present a text in a way in which the writer can view the whole structure (of headings and paragraphs) of the text at once, for instance in the "outline view" mode (cf. 2.1.1). This allows the author to browse through the text, to insert new headings and to easily re-arrange whole chapters and sections as well as paragraphs. While the latter is something more related to my previous point (drafting), the former probably has a further and even more profound effect. As authors have an easy-to-use tool at hand to be aware of the whole structure of their texts, they pay more attention to organising and constructing the text. My hypothesis is that, at least in the longer run, texts written with word-processors will be more systematically and deeply structured and internal redundancies will be less frequent. In addition, the simplicity with which authors can insert footnotes or endnotes (something very cumbersome in the pre-digital phase) has probably contributed to the proliferation of a writing style, which is "aware" of the potentials of two layers of text (the main text and the footnote section). That is, even without hypertext, hyper-elements (cf. 6.1) have become increasingly widespread and have changed already the nature and structure of academic texts.

Layout: Another effect of word-processing software is that it became much easier to produce nicely laid out texts. While previous authors had to wait until their paper arrived as page proof from the publisher, today's authors have tools at hand that can produce nearly the same results as the professional typesetters. And indeed they do. The majority of academic texts have a beautiful layout already in very early phases of text production, that is still well before "version 1.0"[609]. In some aspects, this may impact on the characteristics of the texts themselves. The "look and feel" of a text is related to our apprehension and appraisal of the content. Properly laid out, a text may give the reader – and the author is also his/her own reader during the process of writing (!) – the impression of being finished or almost. Layout also influences our feeling about the length of the whole manuscript or of particular sections and consequently our assessment of the balance of the sections. Most importantly, while some of the possibilities to emphasise words (underlined or **bold** print) were at hand already shortly before the advent of the computer (in electrical typewriters), one is rarely aware that different font sizes, tables etc. only came with digital manuscripts.

For sure, all these changes in the mode of production per se have not fundamentally changed the products, the texts. However, the combination of them has certainly had an impact on how texts look like today and certain effects on structure and layout are visible.

6.4.1.2 Length irrelevant?

> "It is no longer the space for publication, but the attentiveness of the readers which is scarce."
> (Rost 1998b, 2, transl. MN)

Based on the assumption that reading on screen is less convenient than reading on paper, one could argue that digital publications have to be shorter than printed ones. However, the technological potentials sustain the contrary.

[609] This notation is common for numbering software versions. Everything below 1.0, for instance 0.9, is considered a preliminary or "beta" (as opposed to "alpha") version. Users of such pre-versions have to be aware that there may be still errors and malfunctions (so-called "bugs").

First, computer screens are being improved by the minute. They will probably reach a level of quality comparable to paper soon (cf. 2.3.1). Reading on screen for long hours will be as convenient as reading from paper. Second, if the Internet is only used as a fast delivery mechanism, length is again no issue as long as the network has enough bandwidth and capacity to allow for these file sizes to be delivered online. Except for extreme cases of multimedia-loaded publications, even printing on a local printer will be no problem. Third, there are heavily enforced "restrictions of space" in the paper world that basically root in printing costs. Word limits both for books and journal articles guarantee the publishers to keep costs low enough to sustain their business. The emphasis on brevity was encouraged by an "expensive system of limited capacity" (Odlyzko 1994, 21), not by author requirements (Hockey 1997b, 3). By contrast, in E-publishing it does not make much difference whether a journal article is 5000, 8000 or 15000 words long as most of the "typesetting" and formatting is done automatically anyway and hence adds not much to the overall length of time needed to finalise the paper for publication. A duplication in file size from, say 100 to 200 kilobytes is almost not noticeable for the user, neither for the publisher as costs for memory space still drop every year. Fourth, the notion of "too long" is obviously a subjective category. It depends on several factors such as one's interest in the topic or the quality of the structure of the text whether you feel a text is too long or not. Hypertexts are likely to be much more structured than the traditional linear text (cf. 6.4.1.4). In addition, whether one is content with the length of a digital document probably depends on the availability of sophisticated navigation tools etc. (see my analysis of the readability of digital media, below 6.4.3). Consequently, the subjective impression of the length of an academic text in the digital format may perhaps even allow longer texts. Finally, consider that even today hundreds of E-books are being read around the world. Thus, even book-long texts are possible in the digital world.

Hence I come to the conclusion that, at least from the perspective of technology, the length of a text is rather irrelevant in the electronic environment (see also Mueller 1997, 9). There are, for sure, other reasons justifying the enforcement of length limits. "Longish", excessively long papers may be not focussed enough. Shortening often contributes to more precise and concise argumentation. However, as is well known to potential readers, "artificial constraints result in articles too telegraphic to be useful either to experts or to non-experts" (Ginsparg 1996, 5). In any case, this is a question of style, which is always relevant, regardless of any arbitrary word limit. In many instances, the current restrictions lead authors to drop whole parts of an article (which tend to be submitted to another journal with a slightly edited introduction) or to abandon a table or graph. Cyberscience will probably lift these restrictions and may contribute to article lengths more tuned to the actual necessities of the subject dealt with than to the accidental limits included in the arrangements between publishers and editors.

6.4.1.3 Fluidity

Under cyberscience conditions, not only the production process is digitised, but also the published product will increasingly remain in digital form as E-publishing becomes ever more widespread (cf. chapter 7). The only way of amending or adding a printed document is by publishing subsequently an addendum, a rectification, or a comment. That is, however, practised only very seldom. The pure fact that the academic article is now bits and bytes instead of paper and ink has some important repercussions. In sharp contrast to printed matter, which cannot be changed centrally once delivered, digital documents are

not inalterable. Digital files are inherently unstable. The history of European science can be interpreted as the attempt to replace unstable texts by stable ones and, in all existing copies, identical texts. Hence, with E-publishing, "the problem of instability of texts in science comes back" (Stichweh 1989, 49, transl. MN).

I distinguish between the following four[610] basic variants of what may be called the fluidity of the digital medium:

a. The digital document may be moved to another digital location. In some respect, this variant is almost identical to a complete removal (below b). However, it can be seen as an interim problem.
b. The only digital version of a manuscript can be removed even after publication; only privately made copies or printouts may "survive" (similar to the state of affairs in the pre-Gutenberg area).
c. Different digital versions of an academic document may circulate. They may either appear to be the same but differ in parts, or be tagged explicitly as a new version. Also the one and only digital version of a manuscript may change over time as either little typographical errors are corrected or even far-reaching amendments or corrections are made.
d. As regards large hypertext structures (cf. my scenarios in 6.3), another form of fluidity can be described: such structures are dynamic as they constantly grow with new modules being added and new links being set. New paths arise on the fly. No one reader reads the same text modules nor *in* the same structure.

In the following, these four variants of fluidity will be analysed and assessed.

ad a. – Move to another location

URLs are often changed without notice and without re-direction. Above all, this concerns much of the primary sources found on the Internet (official documents, statues, news items etc.). Even the URLs of academic journal articles change from time to time in such a way, although it should be in the interest of the publisher/editor to advise the readers of the new location (at least in the form of an automatic forwarding from the old to the new address).[611]

This is no technical problem: the URLs do not change by themselves. Basically, this is an organisational issue. There are a number of initiatives under way to solve the problem (see 2.1.2). The basic idea is to attribute a unique number to each item and have a central server storing the relation between the numbers and the current addresses. The latter have to be kept up-to-date on the basis of some organisational model involving either the originator of the file, the webmaster of the storing server, and/or the libraries. Updating of references in endnotes to keep track of changing URLs etc., i.e. citation timeliness, is an insurmountable problem for conventional journals (this concedes even the E-journal critic Raney 1998, 4). But E-publications of the future will have time-resistant citations. My expectation is that this first variant of fluidity will not be there forever as technology evolves and people are ever more aware of this problem.

[610] Graham (quoted in Armstrong/Lonsdale 1998, 47) proposes to distinguish three types of document change in the E-world, namely accidental (loss of final version, copying errors), well-intentioned (updates, restructuring) and fraudulent (damaging the work of others, covering one's tracks etc.).

[611] For sure, there are structural limits to our hope for the self-interest of the publisher or editor. Both can go out of business or simply lose interest altogether. This issue will be discussed under the label of "who is responsible for archiving digital media" (7.4.2).

ad b. – Complete removal

Removal may be accidentally or deliberately. In the first case, we may also speak of the question of durability, that is the length of time that the article is available for communicative transactions. The accident may consist of copying errors or the destruction of the memory medium. In principle, accidental removal is avoidable and a question of technical backup measures and archiving of E-documents since "there are no technological reasons for their life to be limited" (Treloar 1996, 142). I shall come back to the archiving issue due to the impermanence of the medium and technical errors in 7.3.3 (for a technical overview, see already 2.5).

The second case is trickier. If a digital document is deleted deliberately this can have two fundamentally different causes. Either it is a fraudulent act by a third party or it is a deliberate decision by the author or editor. If someone intends to damage the work of others, this is a matter of data security measures, both on a technical and an organisational level. Here we may include cases where the author wants to cover his/her tracks, that is without being allowed to do so or even against an explicit decision to the contrary (e.g. if s/he regrets the decision to publish something poorly reasoned and wants it removed from his/her personal record). In principle, this form of removal can (and should) be avoided by regulating access.[612] In addition, we would need to find convincing ways to control the integrity and authenticity of all versions of a text. Appropriate technical and administrative measures should assure that readers have confidence that the version they read "has not been tampered with" (Frankel et al. 1999, 5). This could be done by the technology of digital watermarks (cf. 2.3.4.2) or by trusted downloading stations such as official journal websites.

If it is well intentioned (and in good faith), we are in the face of an almost completely new situation, which has to be carefully assessed and needs a decision by the academic community. In the printed world, a similar situation is only known in the case of defamation or similar procedures where a court may issue a mandate to withdraw a printed book or article from the market. However, such mandates are more of a symbolic nature, as they can never reach 100 per cent as soon as the book was already on the market for even a short period of time. In the academic world, however, the withdrawal of books on those grounds is rather exceptional. The electronic environment now allows for other grounds than the interests of third parties to lead to a wish to remove something already published. We have to ask whether withdrawing should be allowed or whether it is even desirable and under what circumstances.[613]

[612] Technically very knowledgeable persons ("hackers") might probably be in a position to even bypass the most sophisticated access restrictions as they have constantly proven to be able to do so, even with highly protected servers of banks or the military. In this sense, we have a qualitatively new situation as compared to the paper world. (Think of the Nazi regime trying unsuccessfully to destroy completely certain books.) It is, however, not to be expected that this might become a real threat to academia.

[613] Recently, Brumfiel (2002) reported an interesting case with the journal Science. A paper had already been published online and in the queue for print publication, but not yet attributed page numbers for the print edition, when the author was accused of fraud. The editor-in-chief decided to suspend P-publication until the fraud procedure was closed and obviously planned to remove the digital version in case the allegations turned out to be true. This triggered a debate around the question whether the removal would potentially undermine the integrity of the scientific record. But even leaving the paper online (together with a note on its status) is problematic as the online version of Science (ScienceExpress) is no guaranteed and secured web address. At this occasion, the editor of Nature made clear that for this journal, the online publication is the final publication, not the paper version.

At a very early stage, that is if practically no one has seen the paper yet, withdrawal is less of a problem. If, however, the publication has already entered the academic discourse (that is if people have already read it, quoted it or otherwise used it), it seems dubious to simply allow removing it because references to it, perhaps even whole pieces written in opposition to it, would lose their fundament. Even if something was a bad piece of science that triggered a discussion or produced a reaction, the source of the debate should be kept for historical reasons. Regarding the practicalities of enforcing non-withdrawal, however, we need to distinguish between formal publication and self-publication on one's own server. In the latter case, it might be easy to establish such a rule, but difficult to implement it. Authors who want to remove a self-published piece will be able to do so in practice. Only if the site was "spidered" or archived in some form by a third party (cf. 2.5), it may nevertheless be outside the competence of an author to remove the file completely. Without any doubt, this form of fluidity will stay with us and is hardly avoidable.

As regards formal publishing in journals, academia is in a better position to implement a rule. The no-withdrawal-is-allowed rule may serve as an important filter to avoid low-quality submissions because submitters may think twice if they run the risk of having one's poor paper exposed indefinitely (Odlyzko 1994, 28). Alternatively, if a paper is only of historical value because its content is outdated (proven wrong by subsequent research), it may make sense to mark it as outdated. This tag could be used to make it less accessible by removing it from the first "layer" of easily available E-publications and instead storing it in an "archive layer" to be accessed only if explicitly asked for.

In many cases, another route than complete removal would be to store a new version next to the original, or to alter the original piece, see next.

ad c. – Different versions and ex-post alteration

The next variant of fluidity is of a somewhat different nature. Changing the content of a file or releasing new versions can be seen not as a vice (as often done), but also as a virtue, a new opportunity provided by the digital medium.

On the positive side, we should list that feedback and reactions may be allowed to actually have a direct impact and to be taken in account by the author. In contrast to the printed world, an author who is convinced that some of the arguments or data in the paper are erroneous may be given the opportunity to rectify the mistakes after publication. Furthermore, even without initial errors, a paper may become incomplete if not obsolete through subsequent research. In particular authors of survey articles (and their readers) may be happy to have the chance to keep it up-to-date (cf. Dewar 1998, 16f.). On an overall level, a system of academic record keeping that allows for the deletion of errors and that is always up-to-date is appealing because learned inquiry can be seen as a continuum: "(T)he reports of its findings – informal and formal, unrefereed and refereed – are milestones, not gravestones; as such they need only be reliably sign-posted. The discerning hitch-hiker in the Post-Gutenberg-Galaxy can take care of the rest." (Harnad 1990; 1998a)

On the negative side, knowledge stability is partly endangered. If there is always a possibility to rectify and amend a result after publication, there may even be less of an incentive to make the piece "water-tight" in the first place (see already Odlyzko's argument above). In addition, different versions of the same text being circulated (and being quoted) may lead to chaos. Which version is the latest? Which one should be quoted?

Since keeping track needs extra time and can only be done for a limited number of texts, Stichweh expects that "text instability could lead to an increasing differentiation among local sub-cultures and the global science system" (1989, 50, transl. MN).

If we allow for different versions of the same academic publication[614], *version control* becomes an issue. The simplest way would be to give each version a consecutive number and to insert forward links to the latest version, or at least to create a website which gives this information. This would give the reader a chance to look up the latest version before quoting it (or to choose to quote an earlier version).

These solutions may be appropriate as regards informal publishing (posting to E-lists, self-publishing). In a formal publishing environment, defining which version is the "published" version and keeping it unalterable seems, however, important. Indeed, there are already examples of E-journals not only allowing for constant review, but making it a core feature: new versions of the same article are stored with a new version (i.e. publication) number next to the old text.[615] This has consequences for quality check: How far-reaching do the changes have to be in order to make it necessary to re-submit it for review (cf. 8.3.3)? Furthermore, this has consequences for publication records: How do we count new versions of the same article (cf. 8.4.1)?

In other cases, discouraging the release of ever-new versions may be a strategy. Frankel et al., for instance, recommend the distinction between

- Informal posting: "when an author offers his/her work publicly, perhaps by presenting it at a conference, posting it on a personal Web page, forwarding it to an electronic listserv, or simply announcing it during a radio or television interview";
- "First Publication": "the version that will be the basis for claims to priority and submitted for certification" should be fix, notified and the author should commit not to withdraw;
- "Definitive Publication": "Once the First Publication is determined, the process of selection and certification begins (...) [which] is the validation process by which the scientific community identifies work that contributes to the production of useful knowledge" (1999, 4f.).

Many commentators would argue that these "definitive publications" are never to be changed. Further research that builds on and upgrades this version with new data and findings should be considered a new publication. As all new publications, it must enter the system anew, "secure its own bibliographic record, and earn its own place in the scientific literature" (Frankel et al. 1999, 4f.). The journals should be seen as archives of academic knowledge production. Opening archives to revision and update should be handled with care. Otherwise, we would end up with "the indefinitely provisional conclusion, the publication that is a permanent draft" (Tomlins 1998, 145).

There are, however, good arguments not to let the debate end at this point. It could be difficult to strictly adhere to version control and the no-withdrawal-rule with regard to the final draft (Kircz 2001, 5). The correction of a typo or a number, or the addition or deletion of a reference can be important. It would be conceivable to let an article "mature", that is get amended over time without losing the essence of the stability requirement. Three versions of this idea have been put forward:

[614] Also multiple versions of the same module could be stored and inter-linked (Hockey 1997a, 4, for the humanities).

[615] For instance, the "Living Reviews of Relativity" (<Cyberlink=679>).

- Instead of withdrawing and/or making numerous further versions an erratum could be permanently linked to the original (instead of stored, if at all, separately as in paper journals, Kircz 2001, 5). This may not only count for corrections, but also for other types of additions (new data, also additional arguments, comments etc.). The basic idea is that we distinguish between the original (core) text and the add-ons. The core would not be altered and always stay the same except for new links at the margin leading to new parts – like the marginal texts of the medieval glossists.
- Alternatively, E-documents could exist in both a fixed and a current version (Armstrong/Lonsdale 1998, 49). One such example is Bailey's electronic publishing bibliography (Bailey 1995 ff.). This situation is similar to the one in which there is a standard paper and an enhanced online (or CD-ROM) version.
- A third option would be to distinguish, from the outset, between two types of papers: fixed documents (print-like) and continuously updated documents (database-like). Making this clear at the top of the document would be mandatory (Treloar 1996, 141).

These ideas would be compatible with the "fixation" and "persistency" key criteria put forward by Frankel et al. (1999) to determine whether something should be regarded as a publication. A document is "fix" if there is a durable record on some medium and it is "persistent" if it remains in the same form and at the same location, so that it is reliably accessible and retrievable over time. The essential requirement for all schemas, which are less strict with the no-alteration/no-withdrawal rule, would be transparency. By this I mean that the reader should always be in a position to know what type of document (fixed or not) s/he holds in hand and where the other, in particular, the subsequent versions are to be found.

ad d. – Hyperbase dynamics

While the above concern all types of E-publications and in particular all present E-journals, the next variant of fluidity is a speciality of hypertext. We may distinguish two aspects:

(1) As will be discussed in more detail below (6.4.2.2), E-publishing offers the reader more options than traditional publishing. It can be described as a "proto-publishing" form in the sense that a document is accessed in a form that is "potential until the person accessing it makes it 'real' or material in a particular way, be it screen, paper or even sound" (Guedon 1994, 2). This idea could be expanded: what the reader reads did not exist in advance, only the act of downloading, of retrieving information "makes" the text, i.e. if it is being put together from a constantly changing database. Another, or the same reader at another point in time, will read something different.

(2) There is still another specific aspect of the dynamism of hypertext structures. E-documents are no longer the smallest exchangeable entities as a hypertext document may be composed of various modules originating from different sources. Many E-documents and webpages are derived from a variety of dynamic databases (cf. 2.4.3). Therefore, E-documents can change with, for instance, time or outside temperature, or stock market index, or rocket launch date. Such an E-document can nevertheless be "the result of some deep science or engineering advance, and hence be a scientific publication" (Kircz 2001, 4).

Both (1) and (2) lead me to the conclusion that hypertext may be considered a "moving target". This would have important consequences for quoting of hypertexts (cf. 7.2.5)

Summing up our discussion of the fluidity of digital knowledge representation, we may conclude that both sides of the debate should and can be reconciled. Stichweh (1989, 49) is right in that extreme text instability would be hardly compatible with a world-wide-

integrated research system. However, also the rhetorical question of LaPorte et al. (1995) why research communications[616] have to be permanent and why they should not change as new information becomes available has to be pondered deeply. Academia has to cope with the new opportunities of removing, versioning and changing of research publications. It has to carefully implement a system that allows, on the one hand, for corrections and amendments with a view to keeping the academic record free from errors and incomplete arguments. On the other hand, a secure publishing environment has to be established in which references made to each and every paragraph in a publication remain valid and useful (cf. also 11.2.2.2).

6.4.1.4 Modularisation

> "(T)he concept of chapters may disappear in favour of a set of related but bifurcating links".
> (Armstrong/Lonsdale 1998, 19)

Besides the effects of digital text production in general, we have to analyse the intrinsic logic of hypertext and hypermedia structures in order to understand the difference between linear text and hypertext.[617] Only few researchers so far have addressed this issue systematically from the point of view of scholarly texts. One exception is Ingraham (2000)[618] who wrote his analysis from a rhetorical perspective. Speaking of the "electronically mediated scholarly argument", this author points at the changing rhetoric that is at the same time triggered by and needed in the new media. Another important discussion of the issues at stake here is to be found in Kolb (1994) – a text on scholarly rhetoric and philosophy which is itself a hypertext. Finally, the study by Carter (1997) on order and structure in non-sequential essays approaches hypertext from a language studies perspective.

This section covers the effects of modularised text production under a variety of aspects: logic, structure, multidimensionality, minimum publication and multiple use versus citation. Furthermore, the transparency effect (below 1) and the problem of the coherence of argumentation in hyper structures will be discussed (below 2).

Logic: The hypertext technology forces the author to follow a certain logic of structuring his/her findings. Experience shows that it is quite difficult to turn a linear text into a hypertext since, in general, the author of a traditional text does not think in a modular way in the first place.[619] The logic of hypertext implies that the size of information bits may decrease, in particular, as it makes much more sense to set very targeted links, which is not possible in the presence of large modules. Perhaps fewer words and high-

[616] This is their new expression to coin the changed nature of the 'paper' or 'publication', see the introductory quote to this chapter on p. 257.

[617] Note that below, in 6.4.3, I shall further explore the readability of hypertexts.

[618] This interesting paper is, however, not yet formally "published" according to the rules of the E-journal JIME, but still labelled "Pre-print under review"; it is nevertheless available in a stable version at the JIME website, see reference list.

[619] An example from this author's personal experience is a general text on cyberscience (in German), which appeared as a standard linear working paper – although in electronic form only (Nentwich 1999a). It has been, together with a hypermedia specialist, turned into a multimedia feature with hypertext structure (Nentwich/Gagliardi 1999). While the first part of that paper was more apt to be modularised and thus made a somewhat readable hypertext, its second part could not be split into smaller chunks without losing the sense of the original argumentation. This part should have been rewritten (but was not in this case due to time restrictions).

lighting such as bullet points might constitute an appropriate style according to this logic (Armstrong/Lonsdale 1998, 33). This pressure towards modularity seems restricting at first, but opens a more ample degree of freedom when it comes to recombining arguments and to referring to them.

Structure: The modular structure is more systematic and explicit, thus modular publications may be 'clearer' than linear ones (Harmsze et al. 1999). In addition, writing hypertexts may trigger more integrated reasoning. Essential features of the hypertext structure are links: the author of a hypertext has to design these links as part of writing. This involves thinking about the relationship between the modules and also thinking in terms of potential readers (making paths "through" the text). In addition, redundancies and repetitions may diminish via links to previous descriptions, examples or arguments. In any case, writing hypertexts would have to be learned first (cf. 6.4.2.1): simply stringing together module after module could otherwise diminish the depth of analysis. Designing the structure of the network of modules is certainly one of the most important tasks of the hypertext author and simply externalising this analytical task by transferring it to the readers would not be enough. Talking about hypertext as a mode of discourse, however, we are still a long way off. The few known examples seem solely to reproduce the quality of layered meaning that all good writing should have – and "without any artistry" (Tomlins 1998, 149).

Multidimensionality: Closely related to the last point, the modular structure allows for enhancing the publication with more material and documentation in a way that opens it up for more dimensions of argumentation and reading. Consequently, "the scholar's role may hover in interesting ways between author, editor, and curator" (Mueller 2000a, 6) and "(c)ollage and montage may be scholarly techniques of the future" (ibid., 4). Furthermore, "strategies of layered presentation" (i.e. hypertext) allow "for the selection of the archive at different levels of detail and sophistication" (ibid.).

Minimum publication: As academic knowledge consists of myriad units (modules) and the connections between them, we may define research as consisting of both defining new modules and making the connections. The two sides do not necessarily have to go together, that is making the connections (as mainly theorists of all fields do) is an activity in its own right. In the event that the academic knowledge would be represented in large explicit hyperbases, Kircz (2001, 10) envisages that "the minimum scientific publication becomes the brilliant insight of a researcher who connects two separate information units by a typed link, without any further business". Hence, the character of what may count as a valid publication may change.

Multiple use vs. citation: If we envisage a modularised E-publishing environment, then one of the most notable features will be multiple use of "modules" instead of citation. The difference would be "that in multiple use, the new author can rely on the completeness and integrity of the original module" (Kircz 2001, 7). If, in a new article, "a description of a machine, the working of a medicine, or a mathematical proof is needed, reference to another work realises a new dimension. Now, we can seamlessly introduce the existing text into the new work." (ibid.) Instead of "re-inventing the wheel", authors would concentrate on what is new in the present paper and only link to these more universal parts. It might well be that in the long run information units are only stored once. "The bottom line is SGML-coded objects that will change their appearance according to the document style demanded by the presentation medium." (Kircz 2001, 8)[620]

[620] For SGML see 2.4.4.1.

Furthermore, modularisation of texts (plus adding multimedia, e.g. in the form of references to external data) has an important effect on transparency of research (1); and on how arguments can be presented convincingly and coherently, see the next section (2).

(1) Transparency

Texts change their character because of the manifold external links: they may become more transparent. We may distinguish two (related) ways how this can be achieved: (a) through what may be called "embedded citing", and (b) through "hypertextured documentation" (a notion coined by Davenport/Cronin 1990, 181f.).

ad (a) *Embedded citing*: The incorporation of quotes in their original contexts or of dynamic data lets the text "transgress" its own boundaries and makes it more embedded in the rest of the literature. In particular, primary or canonical text can be displayed simultaneously so that "the validity or usefulness of the linked text is immediately apparent" (Davenport/Cronin 1990, 186). Even if the reason a reference is given (or not given) by the citing author is not always clear (Kircz/Roosendaal 1996, 8), there seem to be several types of citations: references to supportive or opposing conclusions in the research literature; to empirical evidence (such as case studies or experimental data); to primary documents (legal material, governmental reports etc.); or to the canon of general literature which forms the basis of the current paper. In a sophisticated hypertext environment, the context in which the citing author deems a reference useful could be disclosed via scientific discourse relation links (cf. 6.2.3.2).[621]

ad (b) *Hypertextured* (or multimedia) *documentation*: A related aspect is that more data and additional material and documentation could be delivered together with the core research results. In many cases, it may be of great interest to have a direct look at this supporting information as it forms the basis of the research. For instance, a statistical analysis or the conclusions drawn by a historian could be better evaluated by direct reference to the material used (e.g. Burg et al. 2000, 7). This would again allow other researchers "to check the official or public version of events against the supporting evidence" (Davenport/Cronin 1990, 181f.). Because of the restrictions of publishing space in traditional media, much supporting material could not be made available. This may change in a hypertext environment.[622] A similar vision could be drafted regarding multimedia. Scholarly journals may offer more direct access to and verification of our colleagues' research. For instance,

> "(e)xperiments can be replicated with the reader's own data. Computer programs can be run and their reported efficiency checked on a variety of input. Graphs can be redisplayed with new data. Simulations can be tested. Medical procedures can be demonstrated. Molecules can be rotated. Architectural spaces can be explored. Art work can be studied. Atomic and subatomic particles can be visualized. Sounds can be played and examined graphically. Eventually, even surfaces can be explored tactilely." (Burg et al. 2000, 7)

The vision is that the technology will enable mutual testing, replication, experimentation and verification of research results among scholars. For sure, the technological opportunities (hypertext and multimedia) alone would not do the trick. New "conventions

[621] However, if the function is to please a referee, it is doubtful whether this would be made explicit by the author. Social functions, which would not be fulfilled in the system to be developed, could inhibit its implementation.

[622] On space limitations cf. also 6.4.1.2.

of scientific recording" would have to be established with a view to making controversial material not removed from public scrutiny. Recording as one of the basic aspects of scientific housekeeping and accountability would have to be introduced into the standard routines and methodologies. Furthermore, "as 400 years of scholars have evolved critical strategies for dealing with print-based evidence, we must evolve similar strategies for non-printable sources" (Ingraham 2000, 11).

Many argue, however, that transparency is diminished because the increased availability of information (publications, data, additional documentation etc.) may lead to even more complexity. While this may be true in the short run, special software and services (filtering agents, specialised databases, meta-data etc.) could be able to solve this problem soon (cf. 2.2.2).

When confronted with the idea that transparency might be enhanced in the digital age, my interviewees gave no uniform answers. The assessments differed according to discipline: rather "no" in philosophy and sociology; "yes" in political science, history, language studies and law, and to some extent also in biology and papyrology. As disclosure of information is also part of the disciplinary culture in some fields and as there is not much experience with E-publishing yet, it may well be that the transparency effect will only occur in the longer run. St. Laurent (1992) assumes that those (historians) opposed to this idea of opening up the archives "have something to hide", that they would "fear widespread distribution of the information they 'protect' in any but the most shrouded form". It is certainly true that building an archive or data pool by lavishly collecting the data from various sources or by transcribing, translating and ordering old manuscripts is an achievement in itself deserving some protection. However, in an electronic environment one can think of a number of safeguards preventing possible misuse of such an archive while at the same time serving the understandable claim of the scientific community to be able to evaluate the conclusions drawn before basing further work on them.

(2) Coherence of arguments?

Authors and readers are used to presenting an argument or a "chain" of arguments in linear fashion. Hypertext possibly splits single arguments into pieces (modules) and divides the "chains". Whether this is helpful, neutral or contra-productive is difficult to answer. What about arguing in hypertext? A number of authors have raised doubts whether hypertext is able to deliver convincing arguments or explanations and see its potential in an immense "gateway to resources" and "descriptive capacity" (Tomlins 1998, 144f.). It is being argued that limiting, compressing, bringing "the vast welter of a complicated problem down to a human and manageable scale" (Givler 2000, 7) is exactly the strength of linear argumentation (similarly Landow 1992, quoted by Atkinson 1993, 203). Atkinson condenses this argument in the following sentences:

> "(I)f we push intertextuality far enough, if we take it upon ourselves to explore a large enough range of the previous uses of the signs of which a text is composed, if we broaden the context enough, the reference of those signs and meaning of that text will diminish and dissolve. Meaning is fragile, and the capacity of the network for a theoretically infinite combining and recombining of texts can jeopardize meaning in a fundamental sense." (1993, 211)

In some sense, the hypertext philosophy leads to the impression that everything is linked to everything: even if layered access is provided (that is, if there are special paths that do not give direct access to all modules), at the bottom line, everything is interconnected. While this impression might be a good thing in the first place because this mir-

rors reality, it may not help much in grasping the essentials. Therefore, coherent arguing in a hypertext environment will mean providing guidance to the readers.

It is much too early to give a definitive answer, as we need to gather much more experience with writing and reading hypertexts before we can assess the usefulness of hypertexting. *Further research is needed*, in particular with a view to whether hypertexts may live up to the standard criteria of a text: coherence, clarity, unity, logical order of arguments etc.

For a start, I may already put forward the following points for consideration: It seems that we have to reflect on the notion of 'argument' itself. As a rule of thumb I may say that one argument should go into one module and not be split into various parts (modules). Modules should not be too long; they should not contain too many thoughts at once. Related arguments should be linked together in a string, e.g. with the sequential path or by suggesting a certain reading order to the reader. As a consequence, if we look only at a particular series of arguments, a hypertext reads just as a linear text with the sole difference that the arguments are split into more modules. This does, however, not mean that it is not a hypertext any more since, further on, putting one argument in one module has the advantage that you can refer to each argument separately. This is not only good for commenting, but for using the same argument in different contexts as well.

For sure, (particular) arguments depend on their (argumentative) context. The latter would have to be made explicit to avoid confusion. (It is difficult to predict with so little experience whether this would be a real problem.) If not, the narrative structure of E-publications (in particular of monographs) may change into something more like a "reference work with shorter, possibly non-linear structures" (Armstrong/Lonsdale 1998, 17).

The Amsterdam research on designing useful hypertext structures in science (physics) made clear that "modules grounded in 'common information', and modules with a core of factual information, are easier to create than modules containing the more argumentative novel synthesis of theoretical and experimental insights" (Harmsze/Kircz 1998, 9). One solution could be to separate the first group (common and factual information, i.e. instrumentation, raw data sets etc.) from the interpretation part of a paper and to modularise the latter only superficially into "situation", "methods", "results" and "findings". This means that there might be different types of modules: the ones which do not contain argument and those that do. The latter will be, internally, of a much more linear structure. In addition, lines of arguments will also be represented by paths. In this context, Dicks/Mason (1998, 5.7) favour speaking of "multi-linear lines of enquiry" instead of non-linearity since they assume that "much work will go into the construction of links that enable a coherent line of argumentation to be pursued by the reader".

In sum, modularisation is, at the same time, a virtue as regards making the information and arguments contained in the hypertext more easily and directly accessible (cf. below 6.4.3.1) and a threat to coherent argumentation. There might be, however, various ways to secure coherence in a hypertext environment as well, such as concentrating on useful paths which mirror in some respect linearity and longer, internally coherent modules.

6.4.1.5 Match of object and description

> "The computer can therefore (...) force us to materialize the network of relationships that we are accustomed to holding in the privacy of our imaginations."
> (Toschi 1996, 202)

Reality is complex. Reducing complexity to linear narratives is often not an adequate solution, but nonetheless adopted for lack of alternatives. Hypertext is, by its very nature, a networked approach, which allows a more complex presentation of reality via multiple links between arguments and facts. Through multimedia, academic publications may gain informational content and comprehensibility (Glatthaar 1996, 32). A number of scholars have pondered this potential for their own discipline:

One of the most fundamental problems in writing *history* is writing simultaneity (St. Laurent 1992), something which might be solved in a hypertext environment with multiple windows on the screen. The same seems to be true for *ethnography* where hypermedia has the capacity to allow "a deeper and more rigorous exploration of ethnographic knowledge and a more faithful representation and communication of that depth to readers" (Dicks/Mason 1998, 3.8). The ethnographic object of study can be seen more fully embedded in its social (cultural, economic, political) context, and the analysis may proceed in a more flexible manner through a thicker level of description (ibid., 7.5). For the *humanities* in general, Mueller imagines projects in which authors "immerse themselves in a world, describe or interpret it, and want to communicate to their readers what it feels like to live and move in that world". Here the integration of "visual and verbal materials will offer scholars opportunities for more ample, precise, and nuanced representations of the archives on which their work is based" (2000a, 5f.). There is, however, an underlying assumption that scholars do indeed want to share their resources freely (cf. 6.4.1.4 (1)).

Also the *mathematicians* Grötschel/Lügger argue that E-publications with hypertext structure have the advantage of a better correspondence between the medium and the complex and ramified structure of knowledge (Grötschel/Lügger 1996, 1). Furthermore, there is an argument about how to present best a formal mathematical proof. Obviously, the more formal the presentation is, the harder it is to understand; by contrast, the less rigorously it is presented, the less reliable it is. Therefore Odlyzko (1994, 22, quoting Lamport) suggests an ideal system in which the results are presented in multiple ways, starting possibly with a video of the author lecturing about them (thus, by neglecting details, by intonation and by body language conveying an impression of the real essence of the argument which is hard to acquire from the paper), and going down to a detailed formal proof. In other words, this suggestion is about layering the "paper" and giving multiple paths with different formats to access its contents.[623]

The relationship between the representation and its object is not only touched by the hyper-structure of the publication which may correspond to the tangled reality, but also by new forms of multimedia representations. Just as in the saying that a picture is worth a thousand words, a photograph, a video sequence, a short voice recording, an animated graphic may be more powerful to convey a particular message than pure text. In this context, Treloar reminds us that the "limitations of unassisted print communication in

[623] Also in engineering, this seems to be appealing as Hagler et al. describe the process of putting together an online and CD-ROM multimedia companion to a refereed journal (1998).

the past required the use of elaborate notation systems when discussing non-text information", for instance, traditional scores for music, the International Phonetic Alphabet for language and Labanotation for dance. Now, "hypermedia in journals removes or reduces the need for these and enables e-journals to 'do more' than print" (1999). For sure, these symbolic notation systems would not lose their power of abstraction. They will retain their function for manipulating thoughts on a meta-level. However, there are situations in which the immediate experience is superior to the symbolic communication. In some respects, these non-textual elements may be on an equal footing with the text itself, as Kircz notes:

> "An electronic document is not the electronic version of a traditional paper document with embellishments such as hyperlinks, colour pictures and illustrative animations. (...) in an electronic publication, images, animations and so on cease to be illuminating illustrations to the text, but are now semi-independent knowledge representations that together with the text comprise the scientific argument communicated to peer scientists." (2001, 1; similarly1996, 10)

The non-textual elements can be experienced in a more immediate way and may thus contribute to spontaneous understanding. This is not to argue that researchers could go without text altogether, but the combination of both and not only one of each (text or non-text) may be the answer to some representational challenges.[624]

6.4.2 Effects on authors

Obviously, the new ways of representing knowledge have a number of consequences for the academic author. First of all, new skills are needed (6.4.2.1). But even the concept of authorship as such is in flux (6.4.2.2). Finally, authoring may become a new hybrid activity between analysis and presentation (6.4.2.3).

6.4.2.1 New writing – new skills

> "If the reader is going to read in three directions, then the writer is, of course, going to have to learn to write in three directions – a very different notion of writing from that done in the linear print environment."
> (Atkinson 1993, 209)

Writing in a modular way is not comparable to traditional writing. As already mentioned in 5.1, authors will therefore need to acquire new expertise, namely "translating the conceptual and pedagogical characteristics of the text into an appropriate electronic context" (Armstrong/Lonsdale 1998, 25). Some say that, while a formal method of writing that will exploit the special capacities of publication in the network is needed, it should at the same time retain the values and avoid the drawbacks of periodical articles and monographs (Atkinson 1993, 208). Others doubt this arguing that the core activities of reading and writing remain the same (e.g. Mueller 2000b)[625] and that there is no need for a new "hypertext rhetoric" (McHoul/Roe 1996, 6). Indeed, a series of metaphors used to de-

[624] In this context, Mueller inserts a caveat: comparing the cost-driven constraints of the traditional book with the new constraints of scarcity in an electronic environment, he comes to the conclusion that the constraints will be "cognitive and aesthetic: how much is too much?" (2000a, 5)

[625] Note, however, that in his contribution Mueller is not directly treating hypertext.

scribe hypertexts such as browsing, indexing, searching, maps, filters, tours, navigation etc. "constitute a conventional conceptual reading apparatus" (ibid.). Although the technology may be new, it is argued, the approach to it and the relations to it are, nevertheless, conventional. As we have seen in 6.4.1.4 on the effects of modularisation, this is not convincing. While it is true that their metaphors are also used with regard to traditional writing, they become a reality for the hypertext author as "tours" (i.e. paths) have to be designed before they can be used. This is different from the paper world where it is completely up to the reader how s/he searches, browses and reads the text – the author is just offering one single linear path.

While writing for a hypertext environment also begins with some kind of outline, "the outline of the whole work might be expanded in stages – with each stage functioning eventually as a separate text-stratum" (Atkinson 1993, 209). This is compatible with the present author's own experience in writing an academic hypertext.[626]

This is not the place to expand on how prospective hypertext writing may be like. It suffices to say that there is enough evidence that it will be different from what academic authors learned to do. This relates also to the technical side. Just as scholars typically do their own word processing and document production (cf. 5.4.1), "it is likely that they will also do a large part of their own multimedia production in the future" (Burg et al. 2000, 4). Mastering hypertext processing and perhaps animation software etc. will become a basic skill for anyone in academia – just as using a word processor with all its handy layout functionality has become something everyone has to know.

6.4.2.2 Authorship blurred?

Hypertexts may be realised in various organisational frameworks. In case they are produced by a single author or by a group of closely inter-linked authors (as with the traditional book or journal article, cf. scenarios 1 and 4) and if the hypertexts are separated from other hypertexts (and possibly only linked externally such as today via a reference list, such as in my base scenario), then no problems will arise as to the question of authorship. Through hypertext it may even become much easier for a group of authors to collect their work into a single unit (St. Laurent 1992). In my scenarios 2, 3 and 5, however, I envisage a development that challenges these premises: both a well-defined group of authors (a) and the strict separation between the hypertexts (b) would probably not be the case.

ad (a) In particular, we may think of scenarios in which others, non-authors of the original hypertext, may comment on the modules. Furthermore, linking a module to another module with, for instance, counter-arguments or a divergent view changes the hypertext as originally intended by the author(s). The "author" of a link (remember that links may even bear meaning, cf. 6.2.3.2, and may count as "minimum publication", cf. 6.4.1.4) becomes in a sense a co-author of the hypertext since links are an important ele-

[626] "– Writing a hypertext is a completely different experience than writing a linear paper: right from the beginning you always have the structure as a whole in your mind, you think in structures, in possible reading paths, in modules and links. – Starting with a rough division of sections and subsections and letting the whole evolve over time, it seemed best to add possible paths only at a later stage. – I discovered that there are always two levels present in the author's mind: the first is the standard trail or path just as in any linear text; the second is that one is constantly aware of the opportunity to make side-paths, digressions etc. with a view to give examples, definitions and background or more detailed information without having to worry about this main path, the 'red thread', as it were." (Nentwich 2000b, from the module 'Experiences')

ment of a hypertext structure. In addition, hypertexts will often be authored by more than one person, indeed by a huge quantity of authors who all add bits and pieces to the overall text (hyperbase). In addition, groupware not only enables the de-coupling of parts of a text for editing by other persons, it also enables several authors to work even at the same time on the same text – who is then the author of this text?

ad (b) If we think of a steadily growing hyperbase, consisting of many interwoven hypertexts, it may become impossible to delimit the "borders" of a single hypertext since ever more modules may become part of different hypertexts. In a modular environment, a distinction needs to be made "between documents as the smallest units of communication and documents that are built-up from various components" (Kircz 2001, 5). This might, however, be increasingly difficult as hypertext lends itself to the use of components of others.

In most cases, the authorship of particular modules and links (as well as of groups of links, that is, paths), i.e. the basic units of hypertext structures may be fixed (logged and registered). It may, however, become increasingly difficult to attribute authorship to a hypertext as a whole. This is an important question: How can we attribute the overall scientific result, in particular if the scientific achievement lays (more) with the whole and less with the (small) pieces? Do we have to attribute this at all? If not, how can we cope with the established practice and consequences (legal, economic, social, career-wise) of individual attribution of scientific achievements?[627] Burg notes that the potential disappearance of the one author creates fear (1999, 124).

Will the traditional notion of an "author" disappear? Fuller (1998, 127f.) speaks of the "emerging legal persona of the 'infopreneur' [who] seems to owe more to the twelfth-century compiler-encyclopaedist than to the nineteenth-century genius-author." He points at an interesting parallel of the copying and commenting of the scholiasts in the Middle Ages, i.e. in the pre-Gutenberg area, where "(c)redit, when assignable, would typically go to the person who assembled the most interesting array of texts, leaving aside issues of original authorship" (1998, 127). How could we preserve the concept of authorship (Harter 1996b, 5)? Davenport/Cronin's answer that the system should make a clear distinction between primary and subsequent versions of a text is only valid in some circumstances, but not feasible in others. In particular, if modules are produced in a "very" co-operative way and if only the whole is worth attributing authorship, but not the tiny parts, then it would not help much, if each and every researcher is recorded as a co-author (something technically feasible).

As I shall discuss in more depth in the section on reading (6.4.3), reading hypertexts is less passive an activity than in a linear environment. While some, such as McHoul/Roe (1996), argue against the claim that hypertext is blurring the boundaries between reading and authoring, others envisage ultimately a co-authoring role of the reader. S/he should be able to create her/his own trails through the hypermedia environment and have these saved by the programme. By this token, the environment "evolves with time as the reader becomes something akin to a co-author" (Dicks/Mason 1998, 7.2; similarly Davenport/Cronin 1990, 176; Atkinson 1993, 209). One can imagine a hypertext system in which readers may add text. In this case, the original author – the one who "started" the hypertext – would become superfluous (Flusser 1996).[628]

[627] Owen (2000, 9) argues that the constant relevance of authorship attribution is a problem not only in hypertext, but also in Mode-2 production (Gibbons et al. 1994) which is also triggered by the new mixed-mode of dissemination he is describing.

[628] Cf. also 6.4.3.3, seen from the perspective of the reader.

I expect that for some time still, academic authorship will not be put into question as there are important incentives not to do so (in particular related to career and fame). However, if the academic publication system develops in the directions proposed in my scenarios 2 (hyperbases), 3 (hyperdiscussions) and 5 (knowledge bases), it is likely that the notion of authorship will change.[629]

6.4.2.3 Merger of authoring, information seeking and analysis

The emerging hypermedia knowledge representation may also have an effect on research practice.[630] One can distinguish between hypermedia applications which are "a tool for the presentation of knowledge which is already – to some extent – codified and defined" and those which are "an aid for the accumulation of knowledge about a subject that exists only as 'data'" (Dicks/Mason 1998, 5.1). Hypermedia should not be seen as either a tool for analysis, or a tool for presentation, but rather as offering an approach to research practice (ibid., 5.9). In a hypermedia environment, the various phases or activities of research (cf. 1.1) somehow merge. Searching for basic information and adding new knowledge is not only done in the same medium, it can also be done almost simultaneously.

The CYBERLINKS database, for instance, is both the product of an information seeking process, a research tool and a final product (in a novel representational format). This is similar to publishing in hyperbases in the form of combining existing knowledge in the base through linking and adding new modules of analysis. What I called (6.4.1.4) the "minimum publication" (setting one link between existing modules) above would obviously be such a borderline case between authoring, information seeking and analysis.

In political science case study research (Kulchitsky/Lavoie 2000), hypertext is used, first of all, in the information gathering and analysis phases, but the hypertext system also enables publishing the results. This is similar with programmes like ATLAS.TI[631], which are used to managing large amounts of text (as well as audio, video etc.) gathered in the course of social science research. This software allows making multiple hyperlinks between related passages and analysing them on a meta-level. Recombining related text for the purpose of publishing is easy and there is also an interface to the WWW for collaboration with other researchers. Again, knowledge representation and analysis merge.

6.4.3 Readability of digital media

> "Probably it will soon count among the quality criteria of a text as to which degree it is presented in a readable format." (Rost 1996c, transl. MN)

While the main thrust of this section will be devoted to the readability of hypertext and hypermedia, I shall start with a brief look at consequences for and problems with reading traditional (linear) publications in the digital as opposed to the printed format:

- So far, the "look and feel" of digital publications falls short of the printed counterparts. Many argue that many important principles of the physiology of reading have been abandoned which would lead to rather poor layout (e.g. Munyan 1998). These princi-

[629] See 6.5 for an analysis of the likelihood of the scenarios.
[630] There are already examples of scholars using hypertext systems as tools for their research, i.e. as a means to get additional results (e.g. in case study research Kulchitsky/Lavoie 2000).
[631] <Cyberlink=784>.

ples include, in particular, length of page and width of columns; portrait format; perceived and measurable advantages of paper and books over single electronic displays (e.g. manipulation of multiple pages simultaneously). Munyan summarises this view by stating that "there are strong, long-term correlations between human learning and the physical format of the book" and that "this relationship between humans and books exists despite the presence of alternative reading and learning media" (ibid.). This has also to do with what Schilit et al. (1998b, 2) call the "sense of the document": the layout and page impression (orientation etc.) communicates something of the type of document. They argue that this sense of document is often lost online (similarly Grenquist 1998). Some of the E-book companies have acknowledged these requirements and have developed the E-book of the future, which tries to mirror the paper book as much as possible.[632]

- For the time being, HTML publishing does not allow the author (nor the publisher) to control completely how a publication looks at the receiving (the reader's) end. Is this something we want to keep (Okerson/O'Donnell 1995, Conclusion)? Pagination which is a concept not suitable in the electronic world, is also an issue here (Armstrong/Lonsdale 1998, 19). One increasingly widespread solution are proprietary formats such as ADOBE's PDF (cf. 2.4.4.1).
- A third aspect of readability concerns the quality of output now available from the digital journal databases, such as JSTOR[633]. What the reader gets is generally superior to a copy made from the original (Guthrie 1997, 3). This is because, inter alia, there are no ugly black margins or blurred areas (due to thick bindings) in printouts as opposed to copies made with a copier.

When it comes to reading E-publications in the form of hypertexts, further aspects of readability come into play, which have less to do with layout and physical properties of the paper, but with intrinsic qualities of hypertexts. Academics are used to reading linear text, (strong) hypertexts are different. To begin with, hypertexts are only usable in an electronic environment: printouts on paper lose most of the new and essential features of hypertexts. In many respects, printed hypertexts are an anachronism and hence are much less readable than both a linear printed text and any electronic version.[634] In particular, following links along a path may be very cumbersome and reading the modules in the sequential order of the printout may disguise its content and confuse the reader. In addition, any multimedia features included in the hypertext are useless in the printed world. Consequently, on-screen-reading is a pre-condition of reading hypertexts. Given the state of screen technology, this is in itself a problem (but may be solved soon, cf. 2.3.1). Here, I shall, however, concentrate on reading as such, regardless of the interface.

A basic problem I face in this section (in particular and the whole chapter in general) is that the type of reading and the advantages and disadvantages of hypertext reading can hardly be analysed without concrete scientific hypertexts. "The advantages of the 'new media', the promise of an 'added value' (...) can only be demonstrated by an empirical comparison of media." (Riehm 1996, 14) Unfortunately, there are hardly any such hypertexts available. Therefore, much of what has been said about hypertext reading is ei-

[632] See the overview on eBook.org (<Cyberlink=25>). Some of the E-book devices feature even a leather cover!

[633] <Cyberlink=322>.

[634] With a view to literary texts, the hypertext metaphor seems to offer new opportunities and forms of writing (see e.g. Burke 2001).

ther based on very limited empirical accounts, or rather theoretical and speculative, or it relates to weak hypertexts such as most WWW pages. I shall nevertheless look into the available evidence and the arguments put forward in the literature with a view to drafting a picture of what lays before us. This section starts with an account of potential advantages of reading hypertexts as opposed to linear texts (6.4.3.1) followed by an analysis of the known problems of hypertexts (6.4.3.2). A discussion of the argument that hypertext will trigger new reading habits and eventually lead to an "empowerment" of the reader (6.4.3.3) will close this section.

6.4.3.1 Potential advantages of hypertext reading

> "(H)ypertext may both ease (movement via indices, tables of content, citations and notes becoming transparently easy) and disrupt the flow of the argument."
> (Ingraham 2000, 10)

The issue of readability of hypertexts raises the question of what "reading" actually is. It has been argued that reading is more than categorising and perception, but "remarking and interpreting of differentiating characteristics on a third level" (Wingert 1996, transl. MN). Keeping this wide definition in mind, two main consequences of reading in a hypertext environment will have to be analysed, namely the effects on information retrieval (as one particular type of reading) (2) and the relationship between the information structure and academic reading habits (1):

(1) Better match of traditional reading habits

It is a truism that academics seldom read articles (not to speak about books) from the first to the last paragraph, that is, literally "linearly". What they do is the following:

> "Instead a particular set of interests will lead a reader to an index, then to the selection of an item in print, then (perhaps) to a graphic, or to a cross-referenced item, back to the index, to a different source text and so on. Each item can be thought of as a node, if need be; and (again, if need be) the encyclopaedia and the internal and external texts to which it leads can be thought of as a web of such nodes." (McHoul/Roe 1996, 9)

McHoul/Roe are right in arguing that, in this respect, linear and hypertext reading are quite similar. However, a linear text has only very few hints as to where to find the particular pieces of information one is searching for (namely the hyper-elements as described in 6.1). In addition, many traditional (linear) texts are also far from being easily readable. An elaborated hypertext, by contrast, *may* provide for much easier and quicker access to that bit. Given the general reading habits in academia, a hypertext may be even better readable (in the sense of accessible) than a standard linear text. In this context, consider the power of layout. When assessing the prospects of hypertext we should not judge solely based on the evidence of experimental hypertexts with their many flaws and shortcomings. The experience of the past decade has taught us that better layout made it much more attractive to read 'hand-made' manuscripts and working papers. The rather clumsy Internet of the 1980s (remember the Gopher and WAIS services) only diffused widely when the graphical user interface (WWW) was developed. This indicates that whether or not something is 'readable' and whether or not one is attracted to read it, also depends on the layout and other features not directly related to the content (cf. also Grenquist 1998). Consequently, future hypertexts are those with which to compare present day articles written according to long standing practices of writing linearly.

(2) Advantages for information retrieval

This can also be seen from an information retrieval point of view. Some hold that "(c)omputers are notoriously poor devices for reading, but they can be terrific tools for getting at the stuff to be read" (Mueller 2000b, 3). However, the explicit labelling of modules and links allows for even better information retrieval, both for the human reader and for knowbots. Highly standardised hypertext modules would include such meta-information for easy retrieval. For example, a 'knowbot' would be able to retrieve all modules containing raw data on a particular issue or all published definitions for a particular notion (cf. 2.2.2.2). Also the human reader may easily ask the computer to present it with such a selection for instantaneous consumption. With a view to increasing "consumability" of the digital information we need tools for navigation through all the information (Kircz 1995, 2). Hypertext may be one way to solve the problem for the reader can selectively locate, retrieve and consult precisely those parts of the published works that are relevant, so that the reader's efficiency is increased (Harmsze et al. 1999).

What electronic text can do better than paper, i.e. full or free text search, approaches the problem of disclosure from the author's point of view (since the language is controlled by the author) whereas controlled keywords and thesauri express the reader's, i.e. the information seeker's point of view. Unfortunately, the latter method necessarily lags behind the research language used. By contrast, a domain-specific information representation structure in hypertext format as proposed by the Amsterdam group would solve both problems: more reader orientation and dynamism of the language used (Kircz/Roosendaal 1996, 8).

6.4.3.2 Known problems of hypertext reading

Together with the obvious advantages of hypermedia, a number of problems should also be mentioned. After all, our cognitive capabilities seem crucial both to design hypertexts which are "better" (in the sense of adding value without losing coherence, orientation and readability) than linear texts, and to assess whether hypertexts may live up to our expectations.

Due to the lack of genuine academic hypertexts, there are not yet many empirical accounts and comparisons between the linear approach and hypertext. First experiences with E-books are partly critical. O'Hara/Sellen (1997) report on an experiment comparing reading paper and on-line documents. Their conclusions do not seem, however, to be generalisable since they compare an established practice with a first try on the basis of insufficient software (as acknowledged by these authors). In addition, the texts were not of scientific nature: accessibility and consistency are more important with regard to scientific texts, something like the literary category "tension" is of less importance. A few earlier empirical studies on hypertext reading (summarised by Riehm 1996, 3f.) were inconclusive and inconsistent.[635] Riehm compared how people reacted upon different forms

[635] While one found out that "(l)inear text organization was superior to the hypertext format", the other came to the opposite result. A third concluded that hypertexts were found to have "no decisive drawbacks, but they also proved to have no impressive advantages". Riehm argues that the inconsistencies were "due to insufficient research concepts, lacking consolidation of the individual research results and a subject (electronic information systems in the broadest sense), which itself is subject to strong changes" (ibid., 4). However, he concluded that, in general, "the belief that 'better' results can be achieved by means of electronic books and hypertexts is not supported by the results available".

of project description (printed and electronic) by means of questionnaires. Among others, the questionnaires also asked whether the readers of the electronic version – though obviously reading the material less systematically (?) – have retained more or less than the print readers. But no relevant difference could be found (1996, 11). Riehm acknowledges, however, that the design of his field study might not be sophisticated enough to produce viable results.

Readability of digital texts will certainly become a distinctive quality criterion. Further research is certainly needed. Two main problems are already known and will be discussed in the following, namely the readability of links (1), and the phenomenon "lost in hyperspace" (2).

(1) The readability of links

Hypertexts consist of modules and links. Both are equally important. The reader should be able to recognise them appropriately. We are already used to simple links, which just lead us to another page in the web: a word which is underscored and/or blue while the rest of the text is neither blue nor underscored, a little graphic which changes its appearance when you go over it with the mouse etc. A particular problem is, however, the readability of meaningful links, i.e. those carrying information about the type of relationship between the modules (cf. 6.2.3.2). Since we have seen that there are many different kinds of possible relations, the visual discrimination among them is a difficult task. The Amsterdam group tried to solve the problem with very small icons of different shapes and colours.[636]

While the readability of traditional academic texts has been gradually improved over the centuries and many conventions as to layout, font etc. have been established over time, this process only lays before us with regard to hypertext. We shall not only need further in-depth research on the scientific reasoning and arguing with a view to establishing all possible types of relations (which will most probably be different from discipline to discipline). In addition, good proposals of visual representations of these relations and links (to avoid "internal" confusion) and some sort of harmonisation to avoid various parallel systems used in different contexts (to avoid "external" confusion) are necessary. This will only develop if there are academics who are willing to take up the challenge to both write and read such hypertexts (i.e. we need incentives, cf. 11.2.2.2).

(2) Serendipity vs. 'lost in hyperspace'

On the one hand, hypertext databases allow for the so-called "serendipity" effect known from scholars, wandering and browsing open stack libraries, who may find relevant information simply by looking around on the shelves where books that they originally sought for stand. Now, the researcher must practice "the art of serendipity at the level of the catalog or database" (Mueller 2000a, 3). A hyperbase supports the serendipity effect not only by offering many links to "related modules and topics", but even by offering search results in a list (which comes close to the "book next on the shelf"). Hence, the serendipity effect is perhaps even stronger in a hypertext environment.[637]

[636] <Cyberlink=774>.

[637] Opposing this view, Geser (1996, 11) notes that "the scanning of computerized data banks has to be guided by selection criteria defined *ex ante*", that is readers would have to know in advance what they were looking for. While this is true on a first level (and I should not deny that search-

On the other hand, the serendipity effect may be so strong we could even wonder whether links to external resources are helpful at all, as they distract the readers from the main piece (Armstrong/Lonsdale 1998, 40). It may be true that when reading a hypertext and following links just as they seem interesting, the reader may soon perceive him/herself to be somewhat "lost". There is certainly a strong centrifugal moment introduced through hyperlinks (Wingert 1996, 123). As just argued, this may be seen as a positive effect (serendipity). However, this may become a problem if the reader does not know any more where exactly s/he "stands" in the hypertext structure, how far it is until the end, how much of the whole is already read and how to get back to the main thrust of the hypertext. This is due to the fact that – in contrast to a linear text printed on paper – the reader may have much fewer hints with regard to the whole. To give just two examples: If you read a printed book, it is immediately obvious how much you have already read of the book simply by estimating (unconsciously) the number of pages before and after the current page just flipped open. If you take a book out of the shelf next to the one you were seeking, it goes without saying that you remain conscious that this is now another book you are browsing through.

There are, however, a number of hypertext solutions to this problem. One possibility is a constantly visible web-view of the whole hypertext, which highlights the present module and perhaps shows read modules and followed links in another colour than those not yet seen. Also a side window with the "reading history" (similar to current Web browsers) with links back may be useful. Another solution is a scrollbar as known from word processing software and Internet browsers indicating your position in the whole. Also special displays counting the number of links and modules visited in relation to the overall number (of each path) could help the reader to orient him/herself. It is certainly possible to design particular navigational aides to ensure "that the various trails are clearly signposted, that pathways followed can be retrieved, that departures embarked upon can be logged, and that the intersections of different trails are made explicit" (Dicks/Mason 1998, 7.3). In addition, external and internal links should be distinct in colour or layout with a view to helping the reader making a conscious decision as regards exiting the presently explored hypertext.

In any case, not getting lost in hyperspace and availing from the serendipity effect pre-supposes new skills.[638]

ing databases is an art to be learned, cf. 5.1), the serendipity effect starts to play at the next level of scanning the results from the first query.

[638] See already Mueller (2000a, 3) on simple digital catalogue researching, cf. also 5.1 and 5.3.1.

6.4.3.3 New reading and "empowerment" of the reader?

> "[N]etworking and such computer applications as hypertext 'democratize' information, and permit unprecedented flexibility in text production and manipulation. (...) That may be in nuce the conundrum, the core challenge, of information services in the early online era."
> (Atkinson 1993, 206)

> "This is what text virtuality is, the possibility of taking part and putting back together again a text according to whatever point of view you wish to use. This is what hypertextuality is all about."
> (Toschi 1996, 203)

Author-orientated strategies of digestion dominate the traditional intellectual culture – it is *production-oriented*. First, it is easy to have access to everything the same author has written; second it is difficult to collect what different authors have contributed to the same topic; third, one is bound to the fixed form in which the author has chosen to write. When using computerised libraries, by contrast, this author-bias "is likely to get de-emphasized or neutralized by cross-cutting *reception-oriented* selection criteria chosen by the different users" (Geser 1996, 9).

This tendency is even reinforced in the world of hypermedia. Given the different structure, layout and additional features of hypermedia, it is not unlikely that reading habits would have to change. A hypothesis would be that reading of texts would be more "jumpy", more selective, more eclectic. New kinds of reading can be envisaged: one can speak of reading in "two temporal directions", namely synchronically through the text as provided by the author but also diachronically back through the citations to which the text refers. The second direction is "a kind of bibliographical reading" which allows the reader to "enter the cited text, and read on both sides, so to speak, of the quotation" (Atkinson 1993, 203). I may add that the metaphor of "reading back" is partially misleading as hypertexts may also have 'forward links', that is to subsequent publications. In short, "three-dimensional textuality" allows reading in three dimensions: first, horizontally or linearly within any level of the publication; second, vertically or hierarchically through the levels of the publication; and third, referentially or "bibliographically" (ibid., 208f.). As a consequence of this multi-dimensionality, hypermedia environments open up a radical possibility as "they can be constructed in such a way as to enable both analysis and presentation in the same medium" (Dicks/Mason 1998, 5.2).[639] In this sense, reading a text is increasingly like exploring it.

An additional aspect is that the hypertext structure could potentially become so rich that the structure is more visible than the text itself (Wingert 1996, 125). This leads to the question whether one indeed still "reads" in such structures (ibid., 127). For sure, technically speaking, one also "reads" inscriptions, advertisements and posters on the street, and reading also takes place if you browse the white pages or a catalogue. However, these types of reading of small bits of information are fundamentally different from reading a book. For instance, in a book there may be redundancy as information is presented repeatedly in slightly different contexts. By contrast, in a database-like structure, repetition will be avoided as much as possible. Reading then becomes more a search-like activity.

[639] Cf. already 6.4.2.3.

While some speak of the "limits of (electronic) epistemic empowerment", defined as the capacity to reason and make decisions by means of computational devices and communication networks (van den Hoven 1999, 345), others argue that hypertext reading even has the potential to empower the reader. The reader may become an "interactive user of knowledge" (Dewar 1998, 24). Interaction could be the new paradigm, where the reader will be more actively involved, and hence be less a passive consumer as s/he actively has to seek out which path to take, which link to click, which modules to read and which not. New models like the layered history E-book (Darnton) would lead to a new kind of reading:

> "Some readers might be satisfied with a study of the upper narrative. Others might also want to read vertically, pursuing certain themes deeper and deeper into the supporting essays and documentation. Still others might navigate in unanticipated directions, seeking connections that suit their own interests or reworking the material into constructions of their own." (Darnton 1999, 9f.)

The reader of a hypertext is given a whole new set of choices about how to read and explore a hypertext. Hypertext is "a chance for individualising, actively selecting and hence reader-driven reading" (Wingert 1996, 118, transl. MN). Both internal and external links make reading a hypertext a completely different experience than reading a traditional text. While the reader of a linear text only has the choice of skipping pages of chapters, or starting with the conclusion instead of the beginning, the text itself is written in a way which explicitly favours reading it in one particular way. This is not so with a hypertext. A hypertext does not only allow making choices for further reading (and immediately putting these decisions into practice), but in some cases even forces the reader to do so. In particular with a view to supporting material and arguments, outside references etc., the reader has to decide whether s/he finds it useful to explore the given choices.

> "Whether readers will want to make such choices is an open question, but it remains clear that hypertext has the potential to give the reader control over many decisions previously made by the author(s). (...) Hypertext doesn't yet obliterate issues of control, but moves them to a different level. The author no longer has direct control over the order in which the reader will approach the materials but maintains enough power to place limits on what materials can be read in what order." (St. Laurent 1992, 5)

In some hypertext systems, the reader is even allowed to participate as author by adding comments and even adding new modules to existing works. In this case, some "controlling editor" (St. Laurent 1992) will be needed. Hypertext has thus the potential for readers to become co-authors, i.e. to make publications a two-way-communication (see also above 6.4.2.2, seen from the perspective of the author). For some (St. Laurent 1992, 5), this is the real goal of hypertext. The idea is that strong interactivity would redefine the reader–author relationship, allowing both to collaborate in the production of the work, rather than requiring the reader to passively absorb the production of the writer (ibid., 6). In any case, the writers lose influence on their readers (Geser 1996, 11).

According to the empowerment thesis, cyberscience "comes to consist fundamentally of defining and redefining parameters, so that the reader does indeed become a writer, creating new texts through new contexts" (Atkinson 1993, 203). In this view, there can be no question that the writer will lose substantial authority and autonomy at the level of reading. Authority over how the reader reads is diminishing, as the standard expectation of linear reading is not valid any more. Autonomy is diminishing as the author is writing in an inter-linked web of modules written by him/herself as well as others, and

cannot even avoid the reader transgressing the virtual borders of the author's hypertext. In other words, while the reader is obviously always in primary control of the reading, hypertext and the network increases this control considerably. This leads to the conclusion that, in a fully networked environment[640], "interaction with the textual history of the subject" could become an integral aspect of both writing and reading (ibid., 209). In a hypertext environment with its open and dynamic structure, the textual history is always present in the forms of comments to and earlier versions of the modules.

Opposing the empowerment thesis, a number of arguments have been put forward. First, all links are set by the author, hence the reader is bound to what paths s/he finds (McHoul/Roe 1996).[641] Second, the writer can never control how a reader will interact with the traditional printed book either. Also linear texts may be read in a non-linear mode (Dicks/Mason 1998, 3.6). In this view, hypermedia environments do not enable a radically new form of communication. One can even turn the table:

> "In fact, one can argue that a computer-based hypertext is more limiting than a written text. With the latter, one can physically 'link' from any word in the text to any word in any other text whereas a reader of a computer-based hypertext can only follow the links created by the author, rendering the reader less free to create their own interpretations." (ibid.)

Technically speaking, this argument is out-dated: in a hyperbase it would be no problem to allow the reader to have implicit links from practically all words of the text to a list of further texts which include these words.[642]

Third, and more fundamentally, hypertext enthusiasts' were criticised for their allegedly simplistic picture of standard (linear) "reading" (e.g. McHoul/Roe 1996, 8). Reading is certainly not "a singular process of any kind" but there are various forms and we cannot prespecify the characteristics, which go to make up reading. McHoul/Roe explain that

> "we would always find them in new and varied combinations, in any actual case of reading regardless of whether the activity takes place inside or outside electronic environments. (W)e will always find, in amongst them, characteristics which we should not want to associate with reading as such but which are crucial to that actual case." (ibid., 9)

I come to the conclusion that the arguments derived from the similarity of reading in the linear and hypertext worlds are not easily discarded. I hold, however, that reading in a hyper-medium will be much more active than in the print medium, as there are more choices. In particular, as soon as there are many hypertexts merging together to large hyperbases (cf. 6.3), the situation would not be comparable to the present at all.

[640] I.e. if practically everything is virtually inter-linked and if a reader never has to change the medium, e.g. go to the library to search for the full-text of a reference in a bibliography; cf. scenario 2 (6.3.2).

[641] Also Mueller (2000b, 1) argues that all the digitisation will not make much difference to the way we read. In his paper, however, the notion of hypertext is not even mentioned once.

[642] For example, the online dictionary LEO (<Cyberlink=767>) "links everything online" (hence the acronym), that is, each word in a list of possible translations of the search word is again linked to the dictionary in such a way that following the link produces a new translation list for this word. The PeP (Perspectives in Electronic Publishing, <Cyberlink=494>) experiment does similar things on the basis of a thesaurus (cf. 2.3.4.3).

6.4.4 Enhanced communication?

The traditional journal articles are about to be replaced by what has been coined "research communications" (LaPorte et al. 1995)[643]. Obviously, communication is such "an essential part of E-publishing (Guedon 1994, 4). Having reviewed the developments, various scenarios and a number of other consequences of the cyber way of knowledge representation, I am now in a position to analyse in more depth the following question: In what respect exactly is the communicative potential of digital publications enhanced vis-à-vis the traditional formats?

I found that one prominent effect of the new medium is that, in relation to the paper-based distribution system, scientific results are made available more widely, faster, more easily and in a better accessible form through world-wide electronic distribution, full-text search, hypertext links, meta-tagging, online commenting, interactivity etc. My hypothesis is that this also increases the communicative potential of academic publications. This potential will be explored in more detail in 6.4.4.1 right below. The research results may be taken up, used and built upon more frequently than is the case so far. This may come in three different forms:

a. *Intra-disciplinary*: other researchers in the same sub-discipline may make more direct references to the work of others when writing their papers since the relevant information is more easily retrievable (all scenarios); furthermore, as argued in my scenario 3, hypertext structures may be used for efficient and more-to-the-point academic debates and hence increasing and improving communication; all this may lead to more "connectivity" of research (cf. 6.4.4.2);
b. *Extra-disciplinary*: scholars from related disciplines or working interdisciplinary may gain better access to results in research areas which are not their own; in this context, also transdisciplinarity (Gibbons et al. 1994) that is the exchange of information between research and development outside and inside academia may be favoured. Hypertext has the potential to "open up a discipline" (Davenport/Cronin 1990, 185);[644] and
c. *Extra-academic*: the public-at-large can more easily access the wealth of academic results (for a discussion of the limits, see below 6.4.4.3).

Before I take a closer look at two of these potential consequences I shall explore the communicative potential of digital knowledge representation.

6.4.4.1 The communicative potential

I have pointed out that academic publications should be seen as part of both knowledge processing (academic communication) and the distribution of knowledge (cf. Figure 1-1). Publishing is, however, traditionally a unidirectional activity, that is the author sends out a message and only seldom receives a direct response. The paper publishing system is not in a position to provide for well functioning "back-channels". The only crook-like remedies are letters to the editor, reply articles and quoting in subsequent articles. This is not only (extremely) asynchronous, but also indirect and mediated. In the paper world,

[643] See introductory quote to this chapter (p. 257).
[644] This aspect is discussed in more detail in 10.2.5.

the author normally[645] does not get notified of any reactions to his/her article, not to speak of the reader who was not initially involved in a debate. References to later contributions cannot be implemented in a paper world. In some sense, "documents are darts" (Sumner/Shum 1997, 2), there are no connections between either the document and its surrounding discourse, or the end product and the process that shapes it. The result often is that "questions go unanswered; confusions go unclarified; criticisms go undefended" (ibid.). In short, paper publishing hardly supports interactive academic communication.

By contrast, E-publishing on the Internet may be different in every of those respects. "Forward links" (cf. 6.2.3.2) can be added. There may be also "review links" (Armstrong/Lonsdale 1998, 35), i.e. to sites containing reviews of the E-book or E-paper. In a fully networked environment[646], the author could not only cite another publication but possibly also tag or annotate it "in such a way that the reader is able to infer the author's evaluation or application of the cited work" (Atkinson 1993, 209). Furthermore, E-publications may become interactive. In this context, interactivity can be defined as "the enhancement of user control and influence over the flow of information and the communication" (Neal 1997, 7). For instance, an online debate forum can be associated with each publication to make it the focal point of multilateral communication processes. This may take the form of a whole threaded discussion list added to the main text (see below). Finally, hypertextual E-publications may be annotated by third parties. We may distinguish three forms:

1. *Personal annotation*, i.e. for the scholar's individual use: Fuller points at the fact that annotation is not such a new thing, indeed. Medieval glosses[647] only disappeared with the printing press and are now being reinvented in the electronic medium (Fuller 1998, 127). As soon as the technology is available, E-publications may be annotated personally.[648] Wiley's INTERSCIENCE portal planned to give registered users the opportunity to personally annotate articles online (Grenier 1998, 3).[649]
2. *Public comment*, i.e. visible to all readers: This may either be during the open peer commentary phase (cf. 8.2.1.1) or thereafter. So far this feature, though implemented in some cases, is not yet used as frequently as one might have expected (e.g. Fisher 1997, 4). As the D^{3}E project[650] shows, better technology in the form of sophisticated interfaces, however, may help to trigger discussion and to change the process of publication into something more discourse-like. The example of JIME, the Journal of Interactive Media in Education,[651] is telling as it has an impressive record of online commenting for each article published based on the D^{3}E software. Obviously, there are disciplinary differences.

Many academics discuss the communicative potential of E-publishing. Some highlight the communicative as opposed to the "diffusionary" function of E-publications. One concludes that the "strong interactive bias of the electronic medium" means that even the

[645] Rare exceptions are book reviews collected by the publisher and forwarded to the book author; also articles of the type "A response to ..." or letters to the editor referring to a previously published article fall in this category.
[646] See above fn. 640.
[647] The marginal notes and comments of generations of readers of hand-written books.
[648] See e.g. Marshall (1997; 1998b; 1998a) and Schilit (1998b); see 2.2.2.1 and 2.3.2.
[649] <Cyberlink=783>; at the time of writing in summer 2002, this service seems, however, not to be available any more.
[650] <Cyberlink=57>, see also 2.4.4.2.
[651] <Cyberlink=236>.

journals would tend to evolve into "something that will ultimately resemble the present workings of an academic seminar more than the relatively hardened document of the printed journal" (Guedon 1994, 6; similar Johnston 1998, 11). Not only the final publication, but even more importantly, the so-called "pre-publication continuum" is highly communicative. This is the "phase of scientific inquiry in which ideas and findings are discussed informally with colleagues (currently in person, by phone and by regular mail)" (Harnad 1993, 8). The findings are then presented more formally in seminars, conferences and symposia, and distributed still more widely in the form of preprints and technical reports that have to undergo various degrees of peer review. This pre-publication continuum seems to be undergoing a revolution:

> "It has now become possible to do all of this in a remarkable new way that is not only incomparably more thorough and systematic in its distribution, potentially global in scale, and almost instantaneous in speed, but so unprecedentedly interactive that it will substantially restructure the pursuit of knowledge." (Harnad 1990, 1)

Given these potentials and examples of early implementations, I can therefore, in sum, fully agree with Davenport/Cronin (1990, 186). To comment on previous acts of citation; to quarrel with the judgement of previous readers; to disagree with the author's evaluation of the text; and to make their own comparisons and critiques would represent a departure from the existing conventions of scholarly communication. In any case, the communicative potential of academic publishing is largely enhanced in the age of cyberscience. We shall see how much it will be used.

6.4.4.2 Enhanced connectivity

> "(R)esearch, once controlled by academia is now placed in the free market of ideas."
> (Abeles 1998, 609)

It seems to be a major problem of research today that the various activities are hardly co-ordinated, rarely directly, sometimes indirectly (e.g. by the research agendas of large research funds). In 4.3.5.4, I concluded that there will be more worldwide co-ordination and synchronisation of research activities due to cyber-means. In 6.4.1.4, I explored the changing character of the academic text, which is becoming more transparent. Here, I shall look at a particular consequence of this for scientific communication.

It is almost impossible to get a complete overview of the state-of-the-art in a sub-discipline because of the sheer amount of research done. This state of affairs leads to redundancies and subjective selection criteria. A world-wide hypertext publishing system (both in the form of scenario 1 and the more advanced forms of scenarios 2 to 5) may at least partly solve the problem and may eventually lead to "international networks for the integration of science" (Bainbridge 1999).[652] The key word is transparency: digital information, if well structured, connected world-wide and extensively described with metadata is more readily accessible (see already 6.4.1.4). Hence finding out the status quo of the academic knowledge in a research field is much easier today. In addition, length of publications is less of a problem (cf. 6.4.1.2) and multimedia can be included.

[652] Bainbridge promoted the idea of an "International Network for Integrated Social Science" (cf. already fn. 511).

More transparent knowledge representation in the form of "embedded citing" and "hypertextured documentation" is bound to have a consequence for research as a whole. I define "connectivity" of research as the fact that one piece of research has a good enough "interface" towards the rest of the relevant research, in other words that it fits well and is related to and embedded in the cumulative knowledge. In this sense connectivity makes research accessible and useful for related (parallel or subsequent) scholarly work. There should be widespread agreement that this is a basic characteristic and requirement with a view to advancing science. Connectivity may be enhanced in an electronic environment. In particular, the quoting of sources becomes easier due to easy access to almost everything written about a subject. Hence aggregated footnotes may become widespread.

In the electronic era, more than ever before, all previous scientific reporting, discussions and controversies will be available as permanent sources for referencing, inspiration and, where needed dismissal. This also means that parts of old works can be easily integrated into new works. This could lead to a period of "general information re-evaluation" (Kircz 1998a, 2f.). One might argue that in a hyperbase environment, texts of other researchers might shrink to chunks of texts, out of context. But this danger is probably less imminent than in the paper world because the "chunks" are still embedded in their original context and may thus be read in this context. Therefore, I believe that hypertexts will lead to greater connectivity.

Scholars' work may be better suited to reach the (possibly expanding) target group (Kling/Covi 1995). One possible emerging target group may be scholars from other disciplines interested in inter/transdisciplinary work.[653] It is, however, also conceivable that this might as well lead to fragmentation of knowledge: specialists working in their sub-sub-specialities might become unable to speak to outsiders and hence unable to communicate their knowledge in a "connectable" manner (cf. 4.3.5.3). By the same token, this might further "school" building, i.e. the creation of circles of researchers interested in the same topic, applying the same methodology and tending to quote only members of this circle while ignoring others.

The interviewees for this study have been asked their opinion on whether full accessibility and disclosure of academic output (transparency) may lead to more connectivity, i.e. to a better interface of research results with other results and, hence, to cumulative knowledge production. Except for the sociologists and the mathematicians and some of the economists, all expected an improvement of the connectivity of knowledge through the Internet. Note, however, that most scholars have no experience at all with hyperbases, as there are only a few restricted prototypes up to now. While acknowledging the potential, many interviewees referred to a changed disciplinary culture as a precondition for more connectivity.

In sum, it seems likely that the attractiveness of the potential of opening up the specialists' fields will be stronger. There is a chance that the digital, "hypertextured" publishing and communication environment may, in the long run, lead to more connectivity of research. In the first place, this effect takes place within disciplines (see also 10.4.5). It may also play a role between disciplines, namely favour interdisciplinarity (cf. 10.2.5).

[653] See above 6.4.4, point (b) extra-disciplinary communication.

6.4.4.3 New doors for the academic ivory tower

Different from standard academic publications (in paper or digital), hypertexts may be used to make the information more easily available to outsiders or lay people. At least in principle (what lies outside the scope of this study is the issue of the necessary capabilities to actually access the available information). There would certainly be important disciplinary differences with regard to the accessibility of the standard modules of academic hypertexts for the uninitiated. It may, however, be conceived that a number of modules in a more journalistic language are written with cross-links to some parts of the academic paper. A starting point would certainly be the executive summaries or abstracts, the introductions and conclusions.

We may thus envisage a layered model of scientific knowledge representation: the most accessible layer would be an article written for the general public, linking to the next layer consisting of the most general parts of the hypertext, i.e. the abstract, introduction and conclusions. Various „deeper" layers giving access to the technical details of such a „paper" would follow the former parts. Within the article, different paths may guide the reader according to his/her specific interest and ex-ante-knowledge.

This is, however, by far no easy task. Believing that jumping from one small hypertext module to the next would mean access to the "knowledge of the world" is underestimating the complexity of agency and theory relevant knowledge and the disciplinary, paradigmatic and organisational splitting of the bodies of knowledge (Fröhlich 1996a, 5). Hence, designing the "lay paths" to academia knowledge and the "lay versions" of the modules will remain a very demanding task for experts, such as science journalists. There are already some first examples in this direction. For instance, the German newsletter Telepolis[654] and the Austrian Internet news channel science.orf[655] both publish relatively short contributions with many direct links to academic websites for further reading.

Transparency as discussed above (cf. 6.4.1.4) also has an extra-academic side. In particular, "(t)he idea of the virtual classroom and its virtual university makes the distinction between inside and outside the university community more difficult to maintain" (Morey et al. 1996, quoted by Dator 1998, 621f.). Hence, transparency vis-à-vis the extra-academic world touches upon the wider issue of the relationship between science and the public since the Internet offers a new interface between these two "worlds". It is not unlikely that scholars are already, or will very soon be aware of this new interface. Awareness of another potential public outside the scientific community may change the way the results are presented and worded.

In addition to formal academic publishing, homepages of research institutes, universities and the like are standard. There are only a very limited number of academic institutions without. Individual researchers are about to follow and many are already online. The contents of these pages, however, differ widely, from very basic descriptions plus only postal address, up to sophisticated websites with lots of information about the institution and its research and other activities. For sure, the main addressees of these scholarly showcases are other researchers. There is a great potential of websites of academic associations to improve efficiency with regard to communication between conferences and with a view to enabling academics "not only to pursue their special interests but also to take advantage of all the new methods that will support their work" (Riggs et al. 1998).

654 <Cyberlink=781>.
655 <Cyberlink=782>.

The users may not easily find homepages alone, even if they contain valuable information, such as pre-prints and the like. Projects are therefore trying to set up disciplinary portals with search-engines to make the wealth of information stored in the dispersed homepages available to all members of the specific scientific community. One well-documented project is to be found in mathematics (Plümer/Schwänzl 1996).

However, scholarly websites also address the public in general. A German study on web presence of universities and other research units shows increasing and specific demands for online information (Lederbogen/Trebbe 1999). For a critique as of mid-1998 and a checklist for improving of scholarly press websites see Regier (1998). Some institutes use the WWW presence as an important part of their marketing activities (see e.g. the profound assessment of this aspect in Nolte 1998). Others come to the conclusion that many academics have not yet realised the full potential of the Internet (Teubener/Zurawski 2000, 10).

My interviewees were asked to react to the statement "The Internet increases the chances that people outside of academia get involved in research. Research will be more reflective of its societal uses." As to whether the "ivory tower" would get new doors and windows, the answers were rather sceptical, with the exception of the experts from the social sciences. All other interviewees were split or rather negative. The general impression from the discussion with the experts is that, so far, it is more about shopping windows than two-way-doors. The shopping windows are, however, rated increasingly important in times of budget crises.

In conclusion, it seems fair to say that the Internet has created a new interface between academia and the public. It is by far not yet fully exploited, but given the astonishingly fast growing numbers of Internet users, academia is already about, and very likely to invest much more in, reaching the interested lay public. In some instances, this might be on a direct path, in others only indirectly via science journalists.

6.4.5 On balance

This chapter assessed the benefits and shortcomings of the hypothesised move from linear to hypertexts and from paper-based to digital media in academia. Reviewing the various arguments presented in this part on the impacts of the new ways to represent academic knowledge, there is not much reason to fear this development since, on balance, it seems to do more good than bad. The development offers the potentials for more communication in the academic world, better connectivity, more transparency, a better match of object and description, enhanced opportunities to cope with the complexity and dynamics of the research topics etc. The known problems of hypertext reading, by contrast, seem resolvable and so are the problems of fluidity and blurring authorship. In the next section, I shall look at the factors influencing the development just summarised.

6.5 Analysis of factors influencing the development

A number of factors will influence whether and how far the digital formats of knowledge representation (E-publishing, hypertext, hypermedia) shall be implemented, and if, how long it will take them "to take over" from traditional linear texts, and in what form this will happen. Applying the "change model" on diffusion of my theoretical framework (cf. 1.2.3) I shall discuss here the specific factors playing a role with regard to digital knowledge representation in three groups of intervening variables: technical and functional (6.5.1), institutional (6.5.2) and actor-related factors (6.5.3).

6.5.1 Technical and functional factors

The following specific technical and functional issues play a role on the path to digital knowledge representation:

Screen technology: Since hypertexts need to be read and multimedia to be experienced in the electronic format and cannot be downloaded to the paper world without losing much of their distinctive features (see already 6.4.3), the availability of excellent screen technology is crucial (cf. 2.3.1).

User friendliness of authoring tools: Good software is essential since all the tiny technical activities involved in making a hypertext distract from the main task, i.e. writing and thinking. It should be mentioned that such editors do not seem to exist yet (cf. 2.4.4.3). Therefore, the potential of E-publishing to include hypertext links, multimedia enhancements etc. still remains largely unrealised, as authoring tools and other techniques that would exploit the possibilities are still in their infancy (Wellman/Minton 1998, 9). Potentially, however, they could certainly be developed. They will help organise a hypertext by assisting in making the right type of links and modules. Furthermore, these tools will also do the meta-tagging and storing in the right place of the net environment.

Network reliability: Collaborative hypertexts writing over the Internet involves access to remote file servers and opening remote files locally for editing. In the framework of the DISKURS experiment I used a MICROSOFT Windows NT network (cf. 0.3.4.2) to try out how this would work with present technology. Even though I had a fast network connection between the various computers involved in the experiment, it took much longer to open, save and edit files that were stored on another computer far away. Although it worked reasonably well and, so far, without a major incident, future versions of collaborative authoring tools should face the reliability problem. This could be done by including safeguard mechanisms such as constantly keeping a local backup copy of each file opened and sophisticated tools to restore the files on the remote file server once there was a network problem leading to loss of data.

Lack of standards: A potential obstacle to the move to hypertext and multimedia may be the incompatibility of the different proprietary formats and even the different brands of XML to store hypertexts and multimedia applications (cf. 2.2.2.1 and 2.4.4.1). In large (global), networked versions of hypertext, intensive traffic is to be expected and therefore complex controls, etiquettes and protocols to ensure proper engagement between texts will be required (Davenport/Cronin 1990, 178).

Access to networks and software: I am not talking here about access to the Internet as such (since this is almost universal by now in academia)[656], but about access to servers which would allow for collaborative hypertext authoring. As this author's own experience in setting up the DISKURS server (cf. 0.3.4.2) shows, this proves to be an important issue. The Internet as it stands now, is organised in a way which makes it rather difficult to set up the necessary infrastructure without running high security risks.

Archiving is not only a technical issue involving standard backup software with a view to ensuring that no module gets lost, but also a highly sophisticated organisational issue. Hypertexts, like all pure-electronic texts, may be constantly changed (cf. 6.4.1.3). If we still had somehow closed and fixed texts (or even hypertexts) which, once published, stay fixed, archiving would be less of a problem. If, however, the system evolves towards dynamic "hyperbases", as I called it in scenario 2 (cf. 6.3.2), in which the various modules will constantly be adapted to new knowledge, to refinement and ongoing discussions, the situation would be different. In this case, a new way of dealing with different versions, concurrent access to modules, logging of authorship etc. would have to be developed.

On the *functional task* level, I find that the types of papers to be written are of importance for whether or not hypertexts are at all an option. This is not only a matter of disciplinary differences since, also in one single discipline, various types of papers may be adequate. In particular, empirical papers, i.e. papers reporting an experiment, a case study or a survey, may be treated differently than more theoretically oriented ones in which the authors argue and try to convince. Obviously, some texts may be more easily turned into or written as a hypertext. Apart from this distinction (empirical vs. theoretical), we may further distinguish between the various text genres, for instance encyclopaedia-like texts (like the "Wissensnavigator": Schmidt 1999), textbooks for students, review articles (reviewing the state-of-the-art on a topic – like Living Reviews[657]), book reviews, newsletter contributions, long articles vs. abstract-type papers etc. These different functions of texts may only result in different types of hypertexts. By contrast, I assume that there is no distinct group of genuine linear texts that could not be turned into hypertexts, in principle, and another group of texts that could. For instance, we may need more "layers" (in the sense of groups of modules which are accessible via different paths at different times) in a textbook because it may be the aim of this book not only to structure the information but also to prepare it for easy digestion and learning.

On the *functional disciplinary* level, I find that while some (sub-)disciplines have already taken up the challenge of hypertexting,[658] others lag behind and may perhaps never join. In 3.4.2.2 I explored a possible relationship between the general visual orientation of a field and the spreading of multimedia applications and found, on the basis of my empirical data, no concluding evidence. However, for logical reasons, the perceived need to represent multi-medial content at all, is certainly a pre-condition.

[656] While in the mid-1990s access to hypertext readers was still limited (McHoul/Roe 1996, 3), the equipment required has, however, become ubiquitous with the WWW.

[657] Below 7.2.4.1.

[658] See for example the well-developed project in physics or the first ideas in sociology or history (St. Laurent 1992); details in 3.3.3.

6.5.2 Institutional factors

On the organisational level, I find *institutional inertia* to play a role: When reviewing Vannevar Bush's early ideas about his hypertext machine MEMEX (Bush 1945), Friedewald argues that Bush, as a technician, did not include institutional constraints in his optimistic vision:

> "Bush's ideas about the changes of the scientific work triggered by MEMEX challenge in fact the functioning of academia which has developed since mid 19th century (...) Such institutions, in particular if they have grown over generations and guarantee their members security and reputation, offer resistance against reform projects which should not to be underestimated – even if they are legitimate." (Friedewald 1998, 183, transl. MN)

A shift to a hypertext culture is, in fact, a major change in the way academics do their jobs. They would have to write and co-operate differently; the traditional journal system would be shattered; etc. Institutional structures based on another paradigm will have an important influence on whether or not this change will take place. For instance, publishers and libraries will have a say. Furthermore, promotion systems would have to change: Whether or not the future of writing academic hypertexts will be bright largely depends on the incentives given to academics to explore these new ways of presenting their results. We may draw an obvious parallel here with the way E-journals become accepted in the academic communities. In many disciplines, there are already journals that only exist in electronic format. Not so long ago, a publication in such a medium would not be rewarded by academic promotion boards. This era seems over by now: E-journals, if seriously refereed, are widely accepted. The standard format of linear texts remains, however, the rule. Except for a few enthusiasts, nobody in academia would invest much effort in writing hypertexts if they were not rated as valuable items on one's publication list (cf. 8.4). As with E-journals, I assume that the more hypertext projects there are, the more academics will join and explore these new opportunities. Hyperbases like E-journals are interactive innovations, which means that the critical mass is relatively high – but once it is reached, diffusion is likely to sustain itself (cf. 1.2.3.5). However, I agree with Atkinson (1993, 210) that the replacement of linear reading and writing by a hypertext will depend upon prearranged, universally accepted conventions.

Among the *cultural* parameters, the shift to new forms of digital knowledge production is obviously closely related to the sustained prestige of paper, and hence shares the destiny of E-publishing as a whole (cf. 7.3).

Furthermore, it seems that it is crucial how well structured the knowledge produced in a field is and whether there is something like a *cumulative production* (cf. 3.4.4.3). If this requirement is fulfilled, hypertext is a more likely option. It then comes as no surprise that a group of theoretical physicists chose experimental physics as the field to try out their ideas about hypertexts while the ideas to initiate a discourse mark-up language for sociology were made in 1996 and still await their realisation as a prototype. This is perhaps a hint that, for at least some disciplines (in the social sciences in particular), hypertext would be a rather revolutionary future since it turns many well established practices upside down (in particular: single-authorship; differentiation/dissociation/distinguishing instead of building up together the "canon" of knowledge etc.).

6.5.3 Agency

In the fields just mentioned above, cyber-entrepreneurs (cf. 3.4.5.1) and activating policy entrepreneurs (cf. 11.1) seem of particular importance. In the absence of such enabling factors, the established cultures may resist and the potentials of cyberscience may as well never be realised.

6.5.4 The likelihood of the scenarios

In conclusion, I look again at the scenarios presented in 6.3 and analyse, on the basis of my presentation and discussion of factors possibly influencing the path to a hypertexted scholarly environment, how likely it is that they will actually take place.

Beginning with my *base scenario*, I find that it is rather conservative and already partly with us. An ever-denser net of cross-linked E-publications is being knotted worldwide. Very soon, more or less all publications will be included in this web. By contrast, the other scenarios are more advanced and obviously more premises have to be fulfilled for them to become reality.

Given, on the one hand, the great popularity of E-list discussions and increasingly of E-conferences (cf. 4.2.2), and on the other hand, the difficulties associated with them (cf. 4.2.2.2), it seems probable that we are quickly approaching *scenario 3*. The difference between the threaded E-list discussions and the envisaged open disciplinary hyperdiscussions is mainly one of the software used to enable it. Threads and hypertextured archives are already standard, software allowing for targeted commenting of particular paragraphs is already with us. Hence, the combination of present features may lie around the corner.

As regards *scenarios 1, 2, 4 and 5*, it seems more difficult to venture predictions. The technical difficulties seem minor and will most likely be resolved in the near future. In particular, one of the main arguments against E-publishing in whatever form, namely that the necessity to read on screen will hamper widespread use, is about to lose much of its – so far undoubted – power due to new technological developments. There will be differences according to the needs and writing cultures in the various academic disciplines. While some will go ahead, others will lag behind. Since there is so little experience in the social sciences with hypertexts, it seems difficult to predict whether or not all academic disciplines will join the train, and when. Apart from the subject-related differences, it seems obvious that – like in many other cases where technology offers new opportunities – it will be less the technical, but rather the cultural and organisational issues that will have a crucial influence on the development. It would probably need a different incentive structure and a nucleus of enthusiasts to further these ideas in each subdiscipline. Ultimately, hypertext publishing and databases "will only flourish if they have the full and active participation, both scientific and financial, of the scientists who use them" (Bourguignon/European Mathematical Society 1999, 114). Not least, this means changes in how researchers perceive their relationship to others and how they co-operate.

Incrementally, it seems, we are coming closer and closer to the scenarios presented here. Hyperlinks have become a daily experience for ever more academics. E-publications are becoming more widespread and the more E-journals and online working pa-

pers etc. there are, the more they will be quoted and cross-linked. Ever more databases spin an ever-denser electronic net between the various publications stored in the Web. In particular, cybrarians are working hard to find a commonly accepted and implemented system of meta-tagging the digital documents with a view to making retrieval much easier. The final consequence of this endeavour may easily be a much more modular structure than today. At the end of the day, we may find ourselves in a situation where we "crossed the Rubicon" without even noticing it: from a primarily linear culture to a hypertext culture.

To sum up, the available and soon much more powerful technology seems to be opening up certain development paths, but other factors, in particular cultural ones, will co-determine whether it will happen. At this point, it seems that given a "favourable" cultural and organisational setting, the hypertext technology could, sooner or later, overtake linear texts and revolutionise the way in which scholars work in most, if not all disciplines.

6.6 Conclusions

> "Neither networks nor hypertext will separately bring about a true revolution – but in combination they are indeed very likely to engender a radical transformation in scholarly information exchange."
> (Atkinson 1993, 211)

In this chapter, I explored how cyberscience is about to affect academic knowledge representation. There can be no doubt that we are already witnessing profound changes. Digitisation of both the production of academic texts and of the texts themselves has taken place already. Weak (that is: simple) hypertexts have by now changed the way we seek information. The WWW and the numerous online databases with which we are confronted in every day academic life are full of hyperlinks, making life much easier. Given the fast evolution of word processing software, it is only a small step from consuming hyper-structures to writing them oneself. My scenarios proposed a number of options how hypertext and hypermedia may change the research environment. These scenarios would, however, go well beyond simple linking of digital texts, but involve profound changes of our perception of knowledge presentation. Strong hypertexts would not only be linear texts linked together in many ways. By contrast, academic authors would be engaged in a new kind of writing and, perhaps, thinking. In addition, multimedia elements may not only "enhance" and "illustrate" texts, but even develop into a form of publication in its own right.

Today, writing and reading genuine academic hypertexts and hypermedia is still in its infancy. Many of the features, as well as writing and reading habits of the present forms of Internet writing, are heavily influenced by people with a technical or scientific background, hence by people with less writing experience due to their training and day-to-day work. We may expect with Nickl (1996) and Harnad (1990) that the Internet involvement of scholars from the humanities and social sciences with their focus on and enormous experience with written texts will make a huge difference. The first beginnings of the exploration of the scientific discourse relations (notably made by a group of physicists) will

be expanded and may, eventually, lead to a new model of knowledge representation in the digital age. It will have to be adapted to the needs and traditions of the various disciplines.

If this vision comes true, if future academic knowledge representation will be rooted in hyperbase structures, if academics will be engaged in consolidating discipline-wide knowledge bases, if scholarly publications become layered and accessible on multiple paths, then academia will be different from today. I have discussed a number of consequences of these scenarios and have come to the conclusion that the character of the academic text would change profoundly. Furthermore, the task of authors would be even more demanding and diversified and readers, by contrast, would more actively explore (and even co-author, in some sense) than passively "read" in a traditional sense. Finally, intra-academic and extra-academic communication could be enhanced, transparency increased and connectivity improved.

My analysis of the likelihood of the scenarios revealed that the technical issues would most likely be resolved as soon as demand increases. Functional factors will play a role with regard to both the speed of adoption and the concrete design. The crucial issues lie in the institutional realm. In particular, it will be a matter of overcoming institutional inertia and of setting the respective incentives.[659]

[659] See 11.3 on policy recommendations.