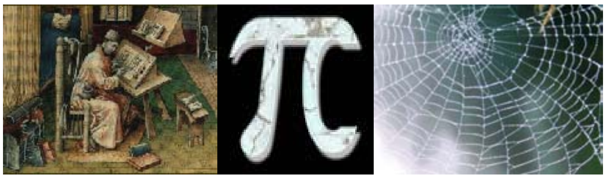
NATIONAL COURSE
IN PHILOSOPHY OF COMPUTER SCIENCE
Proceedings of the Conference



Philosophy of Computing and Informatics Network

Ahonen-Jonnarth Ulla Senior Lecturer PhD Biology Gävle University	**Lager Torbjörn** Professor of General and Computational Linguistics PhD Linguistics Göteborg University
Dodig-Crnkovic Gordana Senior Lecturer PhD Theoretical Physics Mälardalen University	**Lisper Björn** Professor of Computer Engineering PhD Computer Science Mälardalen University
Gustafsson Jan Senior Lecturer PhD Computer Science Mälardalen University	**Nivre Joakim** Professor of Computational Linguistics PhD Linguistics Växjö University
Funk Peter Senior Lecturer (docent) PhD Artificial Intelligence Mälardalen University	**Odelstad Jan** Senior Lecturer (docent) PhD Theoretic Philosophy Gävle University
Correspondent: Gang Liu Deputy Director of Philosophy of Science and Technology Division Institute of Philosophy, Chinese Academy of Social Sciences PhD in Philosophy, Beijing	

Course webpage: http://www.idt.mdh.se/~gdc/04_PI/index.html

SUPPORTED BY
Knowledge Foundation

MÄLARDALEN UNIVERSITY

ACKNOWLEDGMENTS

The financial support from the Swedish Knowledge Foundation is gratefully acknowledged.

The PI-network was essential for organizing the course. Special thanks go to Jan Odelstad and Joakim Nivre as well as Ulla Ahonen-Jonnarth. A great deal of job has been done at home ground at The Department of Computer Scienece and Engineering, Mälardalen University, thanks to Björn Lisper and Jan Gustafsson.

I would like to thank all the excellent lecturers who made this course a success; together with strongly committed participants enthusiastically sharing their knowledge over the borders between the disciplines.

We hope to see the course as well as the network activity develop in the future.

Many thanks to Ylva Boivie for help with organizational issues.

Gordana Dodig-Crnkovic
Course leader
Västerås, Sweden, 2004-06-14

SUPPORTED BY Knowledge Foundation

MÄLARDALEN UNIVERSITY

FOREWORD

The National Course in Philosophy of Computer Science was held during the period January – May 2004, with the following syllabus.

I. PHILOSOPHICAL FOUNDATIONS 22/01 – 23/01

Introductory lecture: What is PI?, Luciano Floridi, Oxford University

Physics as a traditional model of the ideal science for Philosophy of Science, Lars-Göran Johansson, Uppsala University

Philosophical Foundations of Computation, Gordana Dodig-Crnkovic, MDH

II. MTHODOLOGY, MODELLING AND SIMULATION 04/03 – 05/03

Methodological Foundations of Computer Science, Erik Sandewall, Linköping University

Methodological and Philosophical Aspects of Modelling, Kimmo Eriksson, MDH,

Lars-Göran Johansson, Uppsala University

Critical Analysis of Computer Science Methodology, Björn Lisper, Jan Gustafsson, MDH

III EHICAL AND SOCIETAL ASPECTS 13/05

Ethics, Professional Issues, Gordana Dodig-Crnkovic, MDH

Computers in Society - Culture and Art, Gordana Dodig-Crnkovic, MDH

AI and Ethics, Peter Funk

IV MINI CONFERENCE - Presentations of research papers 14/05

Participants from different universities (Blekinge, Dalarna, Mälardalen, Skövde, Uppsala) have taken part in the course. They have presented their research papers at the Mini-conference. Several articles written for the course have already been accepted for international conferences. Here is the list:

Understanding Evolution of Information Systems by Applying the General Definition of Information by Rikard Land is to be published in *Proceedings of 26th International Conference on Information Technology Interfaces (ITI), Cavtat, Croatia-* IEEE*, June 2004*

Correctness Criteria for Models' Validation - A Philosophical Perspective Sandra Ijeoma Irobi is accepted for Models, Simulations and Visualization International Conference (MSV'04)], Las Vegas, Nevada, United States.

Ontological Approach for Modeling Information Systems Imad Eldin Ali Abugessaisa The 4th International Conference on Computer and Information Technology Wuhan, China, 2004 and will be published as IEEE CS.

Early Stages of Vision Might Explain Data to Information Transformation by Baran Çürüklü for Engineering Of Intelligent Systems (EIS 2004)**,** Madeira, Portugal, 2004

Feminist Theory in Computer Science by Christina Björkman, is going to be a part of a PhD thesis.

A few more articles are sent for conferences/journals. An impressive piece of work indeed!

CONTENT

On the Concept of Information in Industrial Control Systems

Stig Larsson
ABB Corporate Research
Forskargränd 8, 721 78 Västerås, Sweden
stig.larsson@mdh.se

Frank Lüders
Mälardalen University
Dept. of Computer Science and Engineering
P.O. Box 883, 721 23 Västerås, Sweden
frank.luders@mdh.se

Abstract

Industrial control systems are computer-based systems that control physical equipment and processes. Most systems used in industry also support interaction with users and other computer systems. This paper describes the different types of data occurring in industrial control systems and discusses the information carried by the data. The concept of information as defined in various fields of science and technology is reviewed, and it is argued that different concepts should be applied to different types of data to provide meaningful and useful interpretations of the data as information. This will also lead to a better understanding among engineers and designers of industrial control systems that the different uses of information put different requirements on the result of their work.

1. Introduction

Industrial control systems are computer-based systems that control physical processes and equipment. Large industrial plants, such as paper mills and oil platforms, are usually controlled by distributed systems, in which several computers perform different control functions while interacting with each other via networks. Interaction furthermore occurs with human users via user interfaces and, of course, with the controlled process via physical interfaces. The purpose of this paper is to identify and discuss *information* occurring in relation to such systems. This includes information carried by the data processed by a system during its operation as well as information arising from the design of a system.

The reminder of this paper is organized as follows. Section 2 describes industrial control systems, including the underlying principles of control theory, the technologies involved in computer-based control, and the main components and structure of the type of systems commonly used in industry. Section 3 reviews various definitions of information, focusing on definitions from English dictionaries and those of the recently identified field of philosophy of information. Section 4 ties together the previous two sections by identifying different types of information in industrial control systems and discussing their relationship to the reviewed definitions. A short conclusion and some ideas for further work are given in Section 5.

2. Industrial control systems

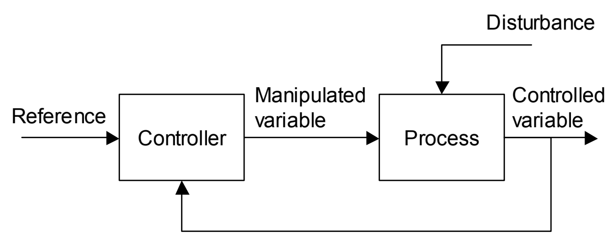


Figure 1. Feedback control system.

Simply put, industrial control systems are systems that control physical equipment and processes. A very important principle used in the design of such systems is that of *feedback control*, which is also known as *closed-loop control* [1]. The principle can be illustrated graphically as in Figure 1. A device called a *controller* measures some output of a physical process, called the *controlled variable*, and compares this to a desired value for the output, called the *reference*. Based on the difference, called the *error*, the controller produces an input to the process, called the *manipulated variable*, intended to drive the process in the desired direction. In this way, the controller can make the process output track a variable reference or keep it close to constant in the presence of a disturbing input to the process. A simple example of a feedback control system is the control of the temperature in a room, where the controlled variable is represented by the measured temperature, the manipulated variable by the flow of hot water through a radiator, and the disturbance by the heat leakage from the room to the outside.

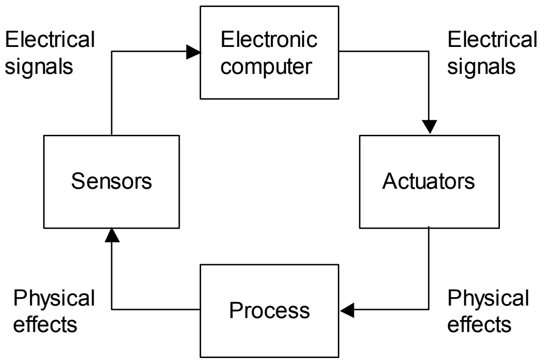


Figure 2. Computer-based control system.

Modern control systems are usually implemented using electronic computers that obtain measures from a physical process via *sensors* and affect the process via *actuators*. This is depicted in Figure 2. A computer typically runs some software that implements the controller part of one or more control loops by, at regular intervals, reading data from sensors and computing and writing data to actuators. For the computer to be able to react to changes in the physical process in a timely manner, the computations it performs are required to produce correct results at correct times. Thus, computer-based control systems belong to the class of *real-time computing systems* [2].

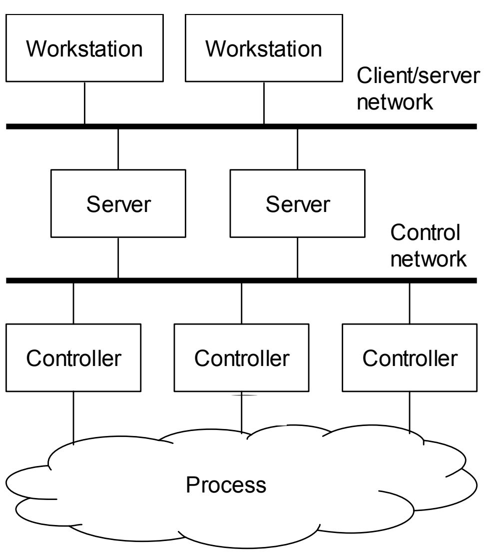


Figure 3. Industrial control system topology.

Industrial control systems are typically built using commercially available *programmable controllers* [3], which have general interfaces for connecting to servers and actuators, and which can be programmed to perform specific control tasks. The systems used to control industrial plants, such as paper mills and oil platforms, are almost always distributed systems where control functions are performed by several such controllers that communicate via some type of network. Typically, the controllers also communicate with other computer systems, such as different types of servers and workstations. Figure 3 is a schematic depiction of a fairly typical topology for such a distributed system.

The purpose of the workstations is to provide different types of user interaction. A typical example is interaction with operators in control rooms that observe and possibly alter the state of the controlled process. Other examples include applications for analysis of process data. Workstations may also be used to run tools that enable engineers to configure and program controllers after these have been installed in the system. An important function of servers is to collect and store process data obtained from the controllers and make this data available to workstations and other servers via the network. In addition, different types of data, such as operator commands and control program updates, must be mediated to the controllers. Software will often be organized as client-server applications, which are executed partially on servers and partially on workstations. The functionality of such distributed control systems at different levels is discussed more thoroughly in [4].

3. Definitions of information

When describing the type of data, information and knowledge in a control system, a precise language is necessary. As discussed by Chemero in [5], different disciplines may interpret the term information in different ways. This section investigates some definitions and views on information. It also contains the classification scheme selected and used in this paper.

Let us first look at some dictionary definitions, all of which are available through the web. This is useful since most persons in contact with industrial control systems (primarily engineers) are most likely to be familiar with these definitions. *Merriam-Webster* [6] has four main entries and several subentries on information. The entries are given here in an abridged version. (1): the communication or reception of knowledge or intelligence; (2a): knowledge obtained from investigation, study, or instruction; (2b): the attribute inherent in and communicated by one or two or more alternative sequences or arrangements of something that produce specific effects; (2c): a signal or character representing data; (2d): a quantitative measure of the content of information; (3): the act of informing against a person; and (4): a formal accusation of a crime.

Web WordNet [7] provides five definitions of information; (1): a message received and understood; (2): a collection of facts from which conclusions may be drawn (statistical data); (3): knowledge acquired through study or experience or instruction; (4): selective information or entropy (which relates to communication theory in which information is a numerical measure of the uncertainty of an outcome); and (5): formal accusation of a crime.

Principia Cybernetica Web [8] repeats some of the definitions but simplifies also the definition into two main uses of the word. The first is related to the definitions attributed to Shannon, "that which reduces uncertainty", and the second to Bateson, "that which changes us".

As a summary of the dictionary definitions we concluded that as expected, they give a general notion of what information is. The different definitions can basically be divided into the two main uses described in [8]; (1): the measure of the quantity of data in a message and (2): obtaining new data that is useful for us. The first interpretation of information relates primarily to the "raw information" that is in focus for *information theory* [9]. The primary aim of information theory is to provide efficient and effective techniques for transferring data.

The second interpretation is more concerned with information as content. Corning [10] describes the two types (1) as being "statistical" and "structural" and (2) as "Control Information". Control information is described as being context depending and user-specific. In our continued discussion it is this second entry that covers our needs best. However, to be able to express the difference between the types of information existing in a control system, a stricter and more granular definition is needed.

A first step towards this definition is expressed by Floridi in [11]. He describes the convergence of a number of analyses into a *general definition of information*, GDI. Here, the content that constitutes information is defined as being *well formed data + meaning.* The description of GDI includes the notions that a set of data is considered as information if it is well formed and meaningful. Floridi notes that GDI is neutral when it comes to the truthfulness of the data. Thus, GDI does not prevent misinformation, disinformation, tautologies etc. to be considered as information. This leads to the *special definition of information*, SDI, which adds truthfulness to the attributes of factual information.

How information can be classified in different types of data is also described in [11]. The defined types are *primary data*, data stored in a database; *metadata*, data about the nature of data; *operational data*, describing how the primary data is used; and *derivative data*, which is data that can be derived from the first three types.

The classification is further elaborated in [12]. Three main categories are distinguished, *semantic information* as defined by the GDI, *instructional information* and *environmental information*. Let us examine the two latter categories a bit further. Instructional information is used to make something happen, and is by nature not false or true. An example of instructional information is the directions from a coach to his or her team before a game. Environmental information can be described as a correlation. This means that the state of one system is reflected in the state of another system. This correlation usually follows some law. However, the correlation does not include any semantic meaning, and this type of information can thus be distinguished from the semantic information. An example of environmental information is the sound on the window indicating that rain is falling outside.

The four types described in [11] are in [12] complemented with *secondary information.* Note also that the terminology now is about information and not data as in [11]. Secondary information is implicit information that may occur through the absence of data, i.e. that something does not happen is also informative.

Let us conclude this section by giving a few classification examples.

- That the signature of a person certifies the identity is environmental information. This is also information of primary type.
- The sentence "the signature of a person is used for identification" is semantic information of the metatype.
- The directions to an operator to close down a nuclear power plant if the alarm goes of are instructional information of primary type.
- The analysis of data from an engine to determine if it is time for maintenance will result in derivative environmental information.
- The information that data from the engine is used for the determination of the need of maintenance is in itself operational semantic information.

- That a person does not answer the phone gives us true semantic secondary information that this person is unavailable, albeit possibly by choice.

- If we, when calling a person, are informed that the person is unable to answer the phone, this is semantic primary information. We may not, however, be able to determine if this information is true or not.

Based on the descriptions and examples above, Table 1 describes what part of the classification scheme is used in Section 4.

Table 1. Taxonomy of information.

Information	Truth-value	Useful types
Semantic	True	Primary Secondary Meta Operational Derivative
Semantic	False	Primary Secondary Meta Operational Derivative
Instructional	*Not applicable*	Primary Secondary Derivative
Environmental	*Not applicable*	Primary Secondary Derivative

4. Information in industrial control systems

This section includes different ways to view the data, content or information in a control system related to the definitions in Sections 2 and 3.

If different views and a classification of the information are used we gain a number of advantages. First, the engineer designing the system will be able to determine the needed properties for different types of information based on the category and type of information considered. An example is the care that must be taken when describing the status of a system for an operator, i.e. since this is semantic information, it must be possible for the operator to determine if the indication of the status is valid or not. This can for example be done through addition of operational data, with an additional indicator that the system is working as expected. Second, as the quality of the data may be determined through the type of data (secondary and derivative data may be considered to be of less obvious nature), more care should be taken when building system that rely on operators taking action based on these types of data. A third advantage is that through the knowledge of different types and categories of information, the engineer may be made aware of the need to include metadata into the system for use in maintenance and when upgrading the system.

In the rest of this section, two different perspectives of the information in industrial control systems are given. The first one is obvious, the run-time perspective. The second one, the engineering aspect, is less apparent, but maybe even more interesting from a philosophical perspective as the use of the information is separated in time from the creation of it.

4.1. Run-time information

During the operation of an industrial control system, data is continuously being processed at several levels, ranging from the electrical signals of sensors and actuators to operator and other user interfaces. From a philosophical viewpoint, it may furthermore be argued that some kind of data or information processing is also performed by the controlled process itself, as well as by human users. This discussion is, however, limited to the information processed by the computer-based control system itself.

Since the data operated on by a computer has both well-defined formats and meaning, it can be considered to carry semantic information, at least as defined by GDI. For instance, the data obtained from a sensor might carry the information "the temperature at measurement point x is y degrees," or, more accurately, "the temperature at measurement point x was, at the time of the measurement, y degrees plus/minus the possible error allowed by the accuracy of the sensor." It furthermore seems reasonable to view this as an example of primary information.

A status flag computed by the controller to indicate whether the measured temperature is below some safety limit or not is an example of derivative semantic information. Data may also be used to carry information of metatype, i.e. information about the information carried by other data. Measurements may, for instance, be augmented with information about the time at which the measurements were made. If a server requests some data from a controller that does not reply within some set time limit, it represents an example of secondary information that the controller or the communication network is not functioning properly. Examples of data carrying operational semantic information are less obvious. One possibility occurs in systems where upload of controller programs is possible during run-time, since such upload would result in data occurring at run-time, carrying information about what information processing a controller is doing.

As for the truth-values of the information carried by the data in a system, two distinct cases can be identified. The first case is illustrated by the above example of temperature measurement. Here, the information can be assumed to be true as long as the sensor is functioning correctly. Thus, it is meaningful to apply SDI in this case, such that correctly functioning equipment yields information and faulty equipment yields *misinformation* (which is not information according to SDI). The other case is exemplified by the data written from a controller to an actuator. The information carried by such data might be "the input voltage of motor *x* shall be *y* volts." It does not seem meaningful to apply SDI to this information, and the fact that it contains "shall be" rather than "is" indicates that it might be more appropriate to view it as instructional information than semantic. This example would most likely represent derivative instructional information, while a new reference value set by an operator would be primary. Secondary instructional information occurs when something like the absence of an expected event informs a component that some action should be taken.

An interesting consequence of the above analysis is that data, which is communicated over a network, may represent instructional information to the sender and semantic information to receiver, due to the possibility of communication errors. Consider, for example, a server sending a new reference value to a controller. From the server's perspective, the data could carry the information "the temperature shall be *y* degrees," the truth-value of which is not subject to interpretation. To the controller, on the other hand, the received data could be considered to carry the information "the desired temperature is *y* degrees," which may or may not be true, depending on whether the data was correctly transferred over the network. Strictly speaking, this observation does not only apply to data sent over a network, but also to data exchanged between different hardware and software components within the same computer.

The above discussions are exclusively based on the definitions on information from the field of philosophy of information as represented by Floridi [11], and not on the dictionary definitions. An important difference between the two classes of definitions is that the latter presupposes a human receiver of the information. Thus, a consequence of using the former class of definitions is that there is no need to distinguish between information on the basis of whether it is observed by humans. For instance, the data obtained from a server and that displayed on a computer screen can carry the same type of information (and indeed, the same information as well), although the formats of the data are considerably different.

4.2. Engineering information

In addition to the information carried by the data produced and consumed by the different parts of an industrial control system during its operation, the system also embodies different types of information that arises from the design, or *engineering*, of the system. The remainder of this section discusses such information, focusing on the information created and used by control system engineers, responsible for building a control system from components such as programmable controllers, sensors and actuators, servers and workstations, communication networks, software applications and tools, etc. On the other hand, the information embodied in the design of those components will not be further discussed. Also not covered is information, such as documentation, which is related to the engineering of control systems but not part of the systems themselves.

An obvious example of engineering information (at least to the authors whose background is in software development) is the programs to be executed by the controllers. As to what type of information these programs represent, several of the definitions seem to apply. For instance, a program is literally a description of what the controller shall do at run-time, and thus it can clearly be viewed as instructional information. In most cases, such programs will be based on other descriptions of the required behavior of controllers, such as sets of differential equations, and should thus be viewed as derived information. Besides being instructional, a program is also a source of information about the data (and thus information) to be processed at run-time. This is clearly information of metatype. The most explicit form of such information is probably declaration of program variables. Since programs consist of data with format and meaning defined by textual and/or graphical programming languages, they can furthermore be seen as representing semantic information as defined by GDI.

Another example of engineering information, occurring on a higher level of a system (according to the schematic of Figure 3), is the process diagrams used by operators to view the state of the controlled process. In modern systems, such diagrams are displayed on computer screens at run-time, using graphical and alphanumerical symbols to indicate process values. They are typically designed using software tools with libraries of symbols representing different types of equipment, pipelines, signal flows, etc. Predecessors of such computer displays are static diagrams on the walls of control rooms, where different types of meters and indicators are also mounted. Common to both forms of diagrams is that they contain information about the controlled process in terms of what equipment it comprises and how this equipment is interconnected. This can clearly be viewed as semantic information as defined by SDI. The truth-value of the information depends on whether the symbols and interconnections on a diagram correspond to the physical process or not. Similarly to control programs, process diagrams are also sources of metainformation about the run-time information.

5. Conclusion

The information in an industrial controls system is used for run-time purposes, but also when building, or engineering the system. Our analysis shows that run-time data can be divided into two classes based on the type of information carried by the data. The first class is most appropriately seen as carrying semantic information as defined by SDI, and the other instructional information. In both cases, we found that all identified sub-types of information may occur. In analyzing to types of engineering artifacts, it was found that each was simultaneously a source of several different types of information.

To be able to discuss the information present in an industrial system, a clear definition of different types of information is necessary. Based on current research and classifications, we propose the use of a subset the taxonomy described by Floridi [12]. The analysis of the information flow in a control system used in industry exposes a number of uses of information. This includes the information at runtime as seen by an operator, the information coming from and sent to the external world, the information built into the control system by the application engineer and others in the design and commissioning of a system, and information that can be obtained from the system to improve the design of future systems. Through the use of the proposed taxonomy, we can identify and focus on relevant properties for each use.

We hope that we through this analysis can encourage engineers and designers of industrial control systems to reflect on the different uses of information. This might serve to make engineers more aware of which data is subject to interpretation with respect to truth-value and may therefore carry either (semantic) information or misinformation. By identifying information as secondary or derivative, engineers may also gain insight about additional potential error sources and take suitable precautions. Furthermore, increased awareness of information types might help designers determine where to add metadata to the systems, e.g. to better facilitate system maintenance.

Future work includes analyzing an existing real-world control system and using the taxonomy to identify and classify the existing information with respect to the different uses. This analysis can be used to determine if the current way of dealing with information fulfills it purpose. Also, this will help in examining if a different way of looking at the uses would change the way that system are designed.

6. References

[1] Åström, K. J. and Wittenmark, B., *Computer Controlled Systems* (3rd Edition), Prentice-Hall, 1996.

[2] Cooling J., *Software Engineering for Real-Time Systems*, Addison-Wesley, 2003.

[3] International Electrotechnical Commission, *Programmable Controllers*, IEC Std. 61131, 2003.

[4] Lüders, F., *Use of Component-Based Software Architectures in Industrial Control Systems*, Technology Licentiate Thesis, Mälardalen University, 2003.

[5] Chemero, A. "Information for Perception and Information Processing", In *Minds and Machines*, Volume 13, Issue 4, 2003.

[6] Merriam-Webster Online, http://www.m-w.com/cgi-bin/dictionary, 2004-02-18.

[7] Web WordNet, http://www.cogsci.princeton.edu/cgi-bin/webwn, 2004-02-21.

[8] Principia Cybernetica Web, http://pespmcl.vub.ac.be, 2004-02-18.

[9] Pierce, J. R., *An Introduction to Information Theory* (2nd Edition), Dover Science Books, 1980.

[10] Corning, ""Control information": The missing element in Norbert Wiener's cybernetic paradigm?", In *Kybernetes: The International Journal of Systems & Cybernetics,* Volume 30, Number 9, 2001.

[11] Floridi, L., "Information", In *The Blackwell Guide to the Philosophy of Computing and Information,* Edited by L. Floridi, Blackwell, 2004.

[12] Floridi, L., "Information", In *Encyclopedia of Science, Technology and Ethics*, Edited by C. Mitcham, Macmillan, forthcoming.

Understanding Evolution of Information Systems by Applying the General Definition of Information

Rikard Land
Mälardalen University
Department of Computer Science and Engineering
PO Box 883, SE-721 23 Västerås, Sweden
+46 21 10 70 35
rikard.land@mdh.se
http://www.idt.mdh.se/~rld

Abstract. *Information Systems are continuously evolved for a very long time. Problems with evolving such systems stem from insufficient or outdated documentation, people no longer being available, different (often old) hardware and software technologies being interconnected, and short-term solutions becoming permanent. Crucial for successful evolution of Information Systems is to understand the data and information of the system.*

This paper argues that some of the fundamental concepts and consequences of the General Definition of Information (GDI), presented by the field of Philosophy of Information, can be a complement to approaches such as "data mining" and "data reverse engineering". By applying GDI it becomes possible for the maintainers of Information Systems to ask important questions about the system that can guide the work in a pragmatic way. GDI can become a useful tool that improves the evolution process.

Keywords. General Definition of Information, Information Systems, Legacy Systems, Philosophy of Information.

1 Introduction

Information Systems are systems whose ultimate purpose is to store and manage information (as opposed to e.g. control systems, which are designed to control physical processes, and manages information only as a means to achieve this). Information systems are long-lived and have to incorporate changing requirements, and thus evolve, often over decades. They become *legacy systems*, with problems such as:

- Documentation is lacking or out of date.
- Very few people have overview over the whole system.
- Different technologies from different eras are mixed.

Still, there is no practical option to start over from scratch. Although there are problems with the system, it represents an enormous effort invested in requirements engineering, designing, implementation, testing, debugging, tuning etc. Evolution and maintenance is therefore often carried out through improving the most urgent parts. To understand requirements and design the system is reverse engineered [1]; to get rid of the oldest and most problematic technologies these parts are ported or migrated [3].

Philosophy of Information is concerned with studying information at the most fundamental level [6]. We will start from the General Definition of Information (GDI) and pursue some of its consequences.

Our contribution is a demonstration of how the GDI can be applied to IS problems to guide some of the activities involved in their evolution.

2 General Definition of Information

Asking what information "is" is probably futile – it is what we define it to be. Dictionary definitions of information typically describe information in terms of *communication*, *data*, *message*, *facts*, *knowledge*, *interpretation*, and *understanding*[1]. According to Floridi, "many analyses have converged on a General Definition of Information (GDI) as a semantic content in terms of *data + meaning*" [6]. Information is, according to GDI, *meaningful*, *well-formed data* – if either the meaningfulness, the well-formedness, or the data is

[1] See e.g. *Merriam-Webster Online*, URL: http://www.m-w.com/cgi-bin/dictionary, *Web WordNet*, URL: http://www.cogsci.princeton.edu/cgi-bin/webwn, *Principia Cybernetica Web*, URL: http://pespmcl.vub.ac.be, 2004-02-18.

lacking, we cannot talk about a piece of information.

GDI leaves a number of aspects of data and information open and does not take a certain standpoint, i.e. GDI is *neutral* with respect to these issues. The openness of these "neutralities" provide a good basis for the analysis and discussion of this paper:

- **Typological Neutrality.** Although information cannot be dataless, GDI does not specify *which types of data constitute information*. Of particular interest is the question whether lack of information means negative information. An example would be whether no answer to the question "how many database entries" should be interpreted as "no entries found", or for example "still processing the query", "stuck in a loop", etc. GDI does not choose an interpretation. Choosing the correct interpretation requires additional information about the data (e.g. meta-data).
- **Taxonomic Neutrality.** Every piece of data is a relational entity. This means that nothing can be regarded as data or information in isolation, but in relation to something else. As an example, a black dot is a black dot only in relation to its background (and the background is a background only in relation to something else). But GDI does not by itself identify either the dot or the background as data, it is neutral with respect to *how data and its relata are identified*.
- **Ontological Neutrality.** GDI rejects the possibility of dataless information: no information without data representation. This has been interpreted as "no information without physical or material implementation" [12], which sounds intuitive when working in the field of computers. Others have interpreted this in another way: that the universe itself at the deepest level is made of information (not matter or energy), exemplified by the phrase "it from bit" [19]. GDI is neutral, however, to the *choice of representation*.
- **Genetic Neutrality.** According to GDI, a piece of information can have semantics not only in the mind of an informee, but also *independent of any informee*.
- **Alethic Neutrality.** According to GDI, information consists of meaningful and well-formed data, independent of whether it is true or false (or contains no truth value at all). That is, GDI does not discuss the *truthfulness of data* (alethic value). This poses a problem for the GDI, and some solutions have been proposed [6]. For our purposes it is enough though to just highlight the issue of truth.

3 Problems Encountered in IS Evolution

For the purpose of this paper, some of the most striking features of an Information System in terms of *information* will be listed. We have chosen to consider the information produced in three development phases, present in all development projects [13,15]. The *requirements* on the system are information. The *design* of the system can be seen as something separate from the implementation – a model of the implementation. For example, the system's architecture, different types of conceptual diagrams are information. As Information Systems are typically divided into (persistent) data and program(s) working on this data, the discussion will consider the design of the data storage separately from the design of the programs. The *implementation* can also be considered information, which again is divided into the implementation of data storage and the programs manipulating data (the programs can be said to embody (part of) the semantics of the stored data).

In line with the taxonomic neutrality of GDI (see above) the relation between each of the above will also be considered as information. This paper will investigate four relations (see Figure 1), because they seem the most natural starting point. First, the relation between *requirements and data design*. That is, why has the data been designed as it is? Second, the relation between *requirements and program design*. That is, why has the programs been designed as they are? Third, the relation between *data design and data implementation*, and fourth, the relation between *program design and program implementation*.

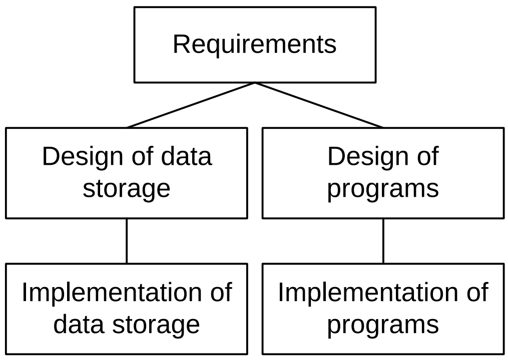


Figure 1: The information considered.

In addition to these relations, one could consider all possible relations between requirements, design, and implementation (i.e. all possible arcs between the boxes in the figure). The relation between requirements and implementation will be touched on, but this is otherwise left as future work.

Before continuing, some notes on the terminology used in the rest of this paper. "Information artifacts" denotes the information produced and the relations between them (that is, the five boxes and the four lines in the drawing). "Developers" denote the people that developed and maintained the system previously. It could be understood as "producers of information artifacts" and includes not only software engineers but also people involved in requirements elicitation. "Maintainers" are the people carrying out the current maintenance and evolution activities, i.e. "consumers of information artifacts".

4 Addressing the Problems

Each of the neutralities will now be applied to the IS problems, one at a time.

4.1 Typological Neutrality

Applied to Information Systems, the *typological neutrality* helps us choosing an interpretation when no information is found. We have chosen the following interpretation: when no information is found, this does not mean that there is no information, but only that we cannot find it.

The opposite seems unlikely. If no information is found, it seems impossible that the information system was built without some notion of requirements and a design (the implementation is of course there, otherwise there is no information system to talk about). There was arguably originally some information, at least in the heads of the developers. This instead leads to the conclusion that the apparent lack of information means that information has been lost. Either it only existed in the heads of the developers, or if it was documented the development organization lacks a document archive. This could lead to either (or both) of the following actions:

- The information should be searched for (if there is the least possibility to find the original developers or old documents).
- An information artifact can be reverse engineered to reconstruct the lacking information. That is, a binary executable can be decompiled, an implementation can be analyzed (even automatically) to produce a higher-level description (i.e. design), and the design artifacts can be analyzed to understand the original requirements on the system.

In addition to this, the organization should learn its lesson and improve its documentation practices, to avoid the same situation in the future.

4.2 Taxonomic Neutrality

The *taxonomic neutrality* means that GDI does not by itself identify a piece of data in relation to something else. In the context of Information Systems, the identification of information as a contrast to something else is very much fixed. For example, the characters constituting the text of documents or the symbols constituting diagrams, as contrasted to the background of paper, has been defined elsewhere and is only *used*. The same is true for the implementation: a language is used with a fixed syntax, which builds on sequences of characters – or if we like, as sequences of bits. But we can widen the question and ask what the information at hand can be contrasted with, in the sense "what is not there but could have been?" And the next question must be "why?"

Of course, one cannot document everything. As a basis, someone writing documentation assumes that the readers will know the language used (e.g. English or UML). But somewhere there is a borderline between what can be assumed, and where it is possible that the reader will misunderstand the intention of the writer. Some terms may be specific to a particular technical domain, or are used in a specific sense in a particular system. Some requirements may have been considered so fundamental that they are never documented as requirements.

Another reason some information is lacking, in the eyes of the maintainer, may be that the documentation practices at the time (or in the company culture) documentation was prepared was different from today's. For example, good documentation practices for architectural descriptions [5,9] (and the very notion of software architecture) is recent, and older systems' architectures may have only been documented very rudimentary [11].

The discussion so far concerns documentation, that is: requirements and design of data storage and programs. For implementation of data storage and programs, matters are different. The information produced could hardly be different (without being erroneous). Perhaps this is because the information in this case is so-called "instructional information" [7], i.e. directions to make something happen, for example a recipe or a sequence of instructions to be executed by a CPU.

4.3 Ontological Neutrality

The *ontological neutrality* states that the information can never be decoupled from its representation. This means that the information was once specified using some data representation. A consequence of this seemingly trivial observation is that the representation chosen possibly affects the actual content of the information. With this in mind, there are numerous questions that should be asked during evolution activities:

Requirements. How were requirements originally represented [10]? Only very informally in the heads of the developers? In a more formal document using natural language? Using some structured notation with natural language (such as a numbered list or a tree structure)? Using some formal language? The answer is itself a piece of information that should be utilized in subsequent evolution activities. Maybe the requirements documentation can be improved by translating it into some more formal form? How did the representation chosen affect the actual requirements – were the requirements focused on functionality or on extra-functional requirements (such as performance, data consistency or robustness)? Focused on data or on behavior?

Programs design. Which languages (textual and graphical) have been used? Flowcharts? Are there architectural descriptions [5]? Is the vocabulary of architectural patterns [4] or design patterns [8] used? The level at which design is made (high abstraction level, such as architecture, or lower level, such as the one modeled with flowcharts) is reflected and affected by the choice of language. This information can be used as a starting point to infer information about the design itself. In case of natural language, do some terms have a specific meaning? Can the choice of language(s) give some clues about the design choices made? The choice of language is partially colored by its popularity at the time the design was made (which may be decades ago) and would not necessarily be the best choice today.

Data design. The same reasoning as for design of programs can be applied to data design, although the languages used would be different: there are e.g. UML [18], so-called "crowfoot" notation, and others.

Programs implementation. How is the implementation represented, i.e. what programming language or languages are used? The choice of language may reflect some conscious decisions based on the information itself, i.e. the program. Especially in newer systems, there are a variety of programming languages to choose from: assembler languages (may indicate a focus on performance), interface definition languages (indicate a component-based approach) [17], logic languages (indicate a rule-based approach) [16], web-based languages (indicates a client-server approach) [14], or procedural languages (may just indicate that a mainstream language was chosen).

Data implementation. The choice of language for implementing data is probably fixed once a database has been determined. Still, the same vendor may offer a variety of languages and technologies for data implementation, and the choice between these reveal some conscious decisions. Standard SQL [2] may have been chosen to achieve portability, or proprietary extensions may have been used, indicate that some vendor specific features were considered more important than portability. For example, non-standard SQL constructs (such as procedures stored in the database server) may have been chosen for performance, security, or data consistency reasons.

Please note that the implementation discussions concern not the implementation on its own but the relation between requirements and implementation – how requirements may have been reflected in the choice of implementation language. Reasoning based on the ontological neutrality is thus one clue (among many) to reverse engineering aiming at understanding the original requirements, in this case mainly the extra-functional requirements. It may even be the case that the decision to use a specific language lays not so much in its technical characteristics as in its political consequences. One example would be choosing a language based on popularity, assuming it will be easy to attract skilled personnel in the future.

Requirements/Data design; Requirements/ Programs design. The relation between requirements and design may have been made explicit in some way. For example, design artifacts such as documents and diagrams may contain references to requirements (in the form of the requirements representation, e.g. the format used to number requirements). Although it is not sure the original developers put effort into making requirements traceable in the design, any clue found is valuable – and it is easier to find these references if they are searched for.

Data design/Data implementation; Programs design/Programs implementation. Source code can easily be searched to find strings used in the design, such as names of higher-level abstractions. Of course, the design description may also explicitly include names of items (such as database tables, column names, interfaces, or classes). Such strings originating from the design could be found in program code (variables, function names etc.) or possibly in comments.

4.4 Genetic Neutrality

According to the *genetic neutrality* of GDI, a piece of information can have semantics independent of any informee. Applied to our discussion, this means that although a maintainer may not understand the information (e.g. the requirements, the design etc.) this does not mean that the information does not have a specific meaning. This may sound as a repetition of the typological neutrality, but there is a difference. The typological neutrality forced us to ask questions about what lack of data mean; the genetic neutrality force us to ask to what extent we understand the data as intended. If the current maintainers do not understand e.g. original design diagrams, these may still have been written in a specific language that were understood at the time of writing.

The genetic neutrality does not, however, state that seemingly unclear texts or diagrams *must* have a meaning that can be discovered if we know the language used. Even if the language used is well known (e.g. English or UML), the information under scrutiny may not conform fully to the language. And it is not uncommon that diagrams are created using an ad-hoc notation with boxes and arrows, without providing a key. A seemingly vague requirement may indeed be vague, even if we know the full semantics of the language used.

Perhaps the genetic neutrality is most useful if interpreted as a procedure: as a maintainer, one should first embrace the attitude that there is information to be retrieved even if it is not understood at once. The language used can provide a key, and to understand the information, one may have to learn the language. This language may be a particular use of natural language (which can be at least partially learned by scrutinizing other documents) or a particular graphical notation (standardized or more ad hoc). The original author, if available, is of course a key person to explain the language used.

4.5 Alethic Neutrality

The *alethic neutrality* highlights the issue of truth: is a certain piece of information true? It seems unlikely that any of the information artifacts would be untruthful on purpose. But there are situations when documents are not trustworthy, which need to be taken into account when using the information contained therein as a basis for evolution activities. Mapped to our information artifacts, it seems unnatural to call requirements, design, implementation, or data untrustworthy by themselves – only when related to each other can they become untrustworthy. This may mean:

Requirements/Programs design; Requirements/Data design. The design might have been insufficient to fulfill the requirements, and therefore the design may be said to be untrustworthy with respect to the requirements. This is particularly common for extra-functional requirements, which are often not analyzed before the system is built (and after it is built it is too late to change the design to fulfill these requirements).

Programs design/Programs implementation; Data design/Data implementation. The implementation may have evolved while the documentation has not. Or the opposite, the implementation never implemented the design fully (due to e.g. time restrictions, which also would explain why the design document was not revised).

As a general principle for software maintainers, the alethic neutrality therefore gives by hand that the information at hand should be met with a sound amount of suspicion. The information should be checked against other information.

5 Discussion and Conclusion

The neutralities of the General Definition of Information have been applied to ten listed information artifacts present in Information Systems, with the aim of discovering important issues to consider while evolving these systems. Most of what have been found is not new; it might rather be seen as old discussions with a new terminology. But there are some things to learn:

- The content and the form of the information are not completely decoupled. So a clue to understand the information is to consider the representation chosen to embed the information: natural language, ad-hoc graphical notations, formal languages, etc. The information reflects the chosen language's strengths and weaknesses.
- Studying the representation used, i.e. what programming languages and design languages were used, may reveal what considerations were important at the time when the information was produced. In particular, this may give clues to the original extra-functional requirements (such as performance or robustness), even if they were not explicit.
- There was once information, even if it is not understood now. Documents and diagrams were written in a language that was understandable for the developer even if they are not clear for the maintainer. Some terms may be left undefined because they were considered trivial by the developer. As a consequence, this attitude leads maintainers to actively search for lost information.

We believe that applying Philosophy of Information to research fields such as that of information system may give birth to new insights. This paper is a first attempt to do this, and there is much left to be done. Future work may include the following:

- We chose three artifacts, from three different development phases (requirements, design, implementation). There are other phases as well to include, e.g. testing. The design phase could be divided into high-level (architectural) and low-level design.

- The actual data stored within the system is also clearly information (hence the term "Information System"). How can the GDI help in understanding and managing this data?
- The other relations between the information listed could be investigated. We are particularly curious about investigating the traceability from implementation and stored data back to requirements.
- Another possible division would be based not on development phases but on architecture, which would typically in an Information System be user interface, business logic, and database.

6 References

[1] Aiken P. H., *Data Reverse Engineering : Slaying the Legacy Dragon*, ISBN 0-07-000748-9, McGraw Hill, 1996.

[2] Bowman J. S., Emerson S. L., and Darnovsky M., *The Practical SQL Handbook: Using SQL Variants* (4th edition), ISBN 0201703092, Pearson Educational, 2001.

[3] Brodie M. L. and Stonebraker M., *Migrating Legacy Systems: Gateways, Interfaces & the Incremental Approach*, Morgan Kaufmann Series in Data Management Systems, ISBN 1558603301, Morgan Kaufmann, 1995.

[4] Bushmann F., Meunier R., Rohnert H., Sommerlad P., and Stal M., *Pattern-Oriented Software Architecture - A System of Patterns*, ISBN 0-471-95869-7, John Wiley & Sons, 1996.

[5] Clements P., Bachmann F., Bass L., Garlan D., Ivers J., Little R., Nord R., and Stafford J., *Documenting Software Architectures: Views and Beyond*, ISBN 0-201-70372-6, Addison-Wesley, 2002.

[6] Floridi L., "Information", in Floridi L. (editor): *The Blackwell Guide to the Philosophy of Computing and Information*, ISBN 0-631-22919-1, Blackwell Publishing Ltd, 2004.

[7] Floridi L., "Information", in Mitcham C. (editor): *Encyclopedia of Science, Technology and Ethics*, Macmillan (forthcoming, see URL: http://www.wolfson.ox.ac.uk/~floridi/pdf/este.pdf), 2004.

[8] Gamma E., Helm R., Johnson R., and Vlissidies J., *Design Patterns - Elements of Reusable Object-Oriented Software*, ISBN 0-201-63361-2, Addison-Wesley, 1995.

[9] IEEE, *IEEE Recommended Practice for Architectural Description of Software-Intensive Systems*, ISBN 0-7381-2518-0, Institute of Electrical and Electronics Engineers, 2000.

[10] Kotonya G. and Sommerville I., *Requirements Engineering: Processes and Techniques*, ISBN 0471972088, John Wiley & Sons, 1998.

[11] Land R., "Applying the IEEE 1471-2000 Recommended Practice to a Software Integration Project", In *Proceedings of International Conference on Software Engineering Research and Practice (SERP'03)*, CSREA Press, 2003.

[12] Landauer R., "Information is Physical", In *Physics Today*, volume 44, pp. 23-29, 1991.

[13] Pfleeger S. L., *Software Engineering, Theory and Practice* (International Edition edition), ISBN 0-13-081272-2, Prentice-Hall, Inc., 1998.

[14] Powell T. A., *Web Design Complete Reference* (2nd edition), ISBN 0072224428, McGraw-Hill Osborne Media, 2002.

[15] Sommerville I., *Software Engineering*, ISBN 0-201-39815-X, Addison-Wesley, 2001.

[16] Sterling L. and Shapiro E., *The Art of Prolog* (2nd edition), ISBN 0262193388, MIT Press, 1994.

[17] Szyperski C., *Component Software - Beyond Object-Oriented Programming* (2nd edition), ISBN 0-201-74572-0, Addison-Wesley, 2002.

[18] UML, *UML Home Page*, URL: http://www.uml.org/, 2003.

[19] Wheeler J.A., "Information, physics, quantum: the search for links.", in Zureck W. (editor): *Complexity, Entropy, and the Physics of Information.*, Addison Wesley, 1990.

Feminist theory in Computer Science

Christina Björkman
Blekinge Institute of Technology
Christina.Bjorkman@bth.se

Abstract

In this article I discuss how feminist theories, particularly within epistemology, can be used within computer science. After an introduction to feminist theory, the article concentrates on issues concerning knowledge in computer science. I approach and explore these questions through a number of themes, which I believe are important to the issues of what *knowledge is produced as well as* how *it is produced and how knowledge is viewed in CS. I discuss for example paradigms and metaphors in computer science, the role of abstractions, the concept of naturalisation and social and cultural aspects in computer science. I also make some brief comments on the Blackwell Guide to the Philosophy of Computing and Information.*

1. Introduction

Can feminist theory be used in computer science (CS)[1]? And if so, can it be used not only for studying and criticising CS, but also for transformation, contributing to the development of the discipline? In this article I want to invite to a dialogue between feminist theory and computer science, what could be the start of a "feminist computer science".

Feminist/gender[2] research concerning computer science has to a large extent focused on issues of gender in relation to computer science, for example the lack of women within computing, and gender equality aspects (see the overview and discussion in [6]). In these studies, CS is often seen as firmly defined, and the underlying perceptions of development and knowledge in CS are seldom brought into focus. In this article my aim is to show how feminist research can be a resource *within* CS, for discussions concerning the discipline.

2. Theoretical starting points

Feminist research represents many theoretical and methodological approaches, and the meaning and focus of the research is different within different disciplines. I will here develop and discuss some of the concepts that I use as starting points.

I base my article and work mainly on feminist epistemologies (see 2.2 below). The website [20] provides a good, brief description of various strands within feminist epistemology, together with suggested readings. In order to ground my discussion, I will start with a very brief look at the development of feminist science critical research.

For a thorough account of feminist/gender research within science and technology, see the work of Lena Trojer, esp [42]. Christina Mörtberg has also developed and discussed these issues [e.g. 30, 31].

For the reader who is interested in different feminisms and their relations to issues of gender and technology in general, there are good discussions of these in [44, 21].

2.1 A brief introduction to feminist science critical research

Around 25 years ago, some feminist researchers started to address science as both knowledge and institution. The starting point was in the claim that science is andro-centric, and has to be changed so that it serves women, in all their diversity, better [35]. It was thus perceived inadequacies and imbalances in established research that motivated a growing feminist critique of science. This feminist science critique developed from issues about women, to realising and focusing on problems concerning how science is constructed and practiced.

This was formulated by Sandra Harding in her groundbreaking book "The Science Question in Feminism" [24]. Harding argued for a shift of focus, from "the woman question in science", by which she means "What is to be done about the situation of

[1] I use the term 'computer science' (CS) in a broad sense, including software engineering and most parts of computer engineering. In all relevant aspects, I use this term as synonymous to the word 'computing'.

[2] Gender research is commonly used as the term in a Swedish context. I choose to use the more common international term 'feminist research' in order to show that the research is not only about gender in itself.

women in science?" [24, p. 9] and towards what is often called "the science question in feminism". A new focus for research thus came to be science itself, its theories, methods and other knowledge processes.

An important feminist science critic is Evelyn Fox Keller. Having herself been an active scientist (within physics and later mathematical biology) she has long and deep experience of the inner workings of science. This makes for a perspective from *both* within science and feminist science studies, making her analyses particularly "to the point". She has contributed to the understanding of the gendering of metaphors [e.g. 18], for example she has discussed in depth issues of what she calls "gender *in* science" [17, p. 86, original italics]. Here, she brings our attention to how gender metaphors work in two directions, or rather in a vicious circle: social expectations influence how we choose representations of nature on the one hand. These representations will then in turn reproduce cultural beliefs and practices.

I have been greatly influenced by the essays in Evelyn Fox Kellers book from 1992: "Secrets of life, Secrets of Death" [16], and in particular the essay "Critical silences in scientific discourse". Here, she discusses how science could be transformed, and points to how representation and form of scientific know-ledge, not only content, is important. Representations will in themselves carry understandings. She argues that we need a more complex understanding of how science works and *for whom* it works.

Feminist science critics commonly prefer scientific models that are interactionist and contextual, over those which are linear, hierarchical, causal or "master" theories [17]. This is in opposition to the Cartesian dualism of mind and body, culture and nature, and the Cartesian project of control and domination. An example of this is that hierarchical structures, eg in biology, can be used to support existing social hierarchies and structures, as well as for domination and control.

Feminist research has always had *transformation* as a prime goal:

"Feminist critique of science [...] is a politically engaged discourse committed to changing both the present organisation of the production of scientific knowledge and the knowledge it produces." [35, p. 20]

The challenge is not to leave the science that the feminist is doubtful about, but to stay and work for changes in the practice of science.

2.2 Feminist epistemologies

Sandra Harding emphasizes the importance of epistemology, or as she phrases it, "concepts of knowers, the world to be known, and the process of knowing" [24, p. 140]. It is particularly important here to note that she does not primarily talk about 'knowledge' as a substantive, but of the activity of knowing, and of knowing subjects. She points out [25] that methodology and epistemology are intertwined with what we do and how we do it, thus underlying all research and knowledge production.

The kinds of feminist epistemologies I build my work on, do not accept the (still strongly prevalent) ideas of science and the scientist as neutral and objective. This is eloquently expressed by Sandra Harding:

"O*bservations are theory-laden, theories are paradigm-laden, and paradigms are culture-laden: hence there are and can be no such things as value-neutral, objective facts.*" [24, p.102]

Harding also points to how science is locally (both in time and place) constituted and the historical contingency of epistemologies:

"Epistemologies are justificatory and methodological strategies that are designed for historically specific purposes." [27, p 281]

My most important influences concerning epistemology come from Donna Haraway. As we have seen above, feminist epistemologies are critical of paradigms of objectivity, and of the neutral and objective observer, what Donna Haraway terms "the God-trick of seeing everything from nowhere" [23, p. 189].

Instead, she develops the concept of *situated knowledge:*

"I am arguing for politics and epistemologies of location, positioning, and situating, where partiality and not universality is the condition of being heard to make rational knowledge claims. These are claims on people's lives; the view from a body, always a complex, contradictory, structuring and structured body, versus the view from above, from nowhere, from simplicity." [23, p. 195]

Situated knowledge is a far-reaching concept, which I understand and use as implying an epistemological standpoint. Thus, *situatedness* refers to conscious epistemological positioning. It is not simply a matter of an individual place or state, it is part of practice and knowledge production, and it means actively taking a stand. And there is no such thing as an innocent position.

So is this relativism? No, the feminist epistemologies that I talk of are not relativist. They attempt to refuse the choice and dichotomy between on the one hand universalism and on the other relativism. Instead, they try to put forward a feminist concept of objectivity.

"*I would like a doctrine of embodied objectivity that accommodates paradoxical and critical feminist science projects: feminist objectivity means quite simply situated knowledges."* [23, p. 188]

Thus, Haraway's alternative to relativism is partial, locatable, situated knowledge.

2.3 Feminist views of knowledge

The common definition of knowledge in mainstream philosophy [e.g. 15] is on the form of: 'S knows that P'. This is only a definition of one type of knowledge, often called propositional knowledge (or sometimes simply theoretical knowledge). This has come to be seen as the only important form of knowledge, at least within western science. What about the knowing subject in this definition? S is not defined here, and thus takes on the form of a universal, disembodied knower, having a view from nowhere in particular. This is the knowledge of the mind, building on the dualisms between mind and body, culture and nature, man and woman etc. Abstract and theoretical thinking is seen as superior (and has been connected to men) while bodily knowing and practical thinking (which has been connected to women) has been seen as inferior. In this view, the body is seen as a hindrance for the 'pure' intelligence of the mind, instead of, as feminist epistemologies claim, an inseparable part of knowledge. Thinking and reasoning are presented as fundamentally mental, of the mind. Skills and tacit knowledge, on the contrary, are seen as a lower form of knowledge than that of the mind. To put it bluntly: Knowledge that is not propositional is not considered knowledge. This has been pointed out by many researchers, for example Sherry Turkle and Seymour Papert:

The formal, propositional way of knowing, has been recognized traditionally as a standard, canonical style. Indeed, philosophical epistemologies has generally taken it as synonymous with knowledge." [43, p. 114]

(Compare this quote with the view of knowledge in for example chapter 17 in [15], see also section 4.1 below.)

Some feminist thinking, as a contrast to the above view, focuses on (women's) material and practical experience of the world, and the kinds of knowledge derived from such practices. This includes the body, and not only the mind. Thus, feminist epistemologies acknowledge (embodied) experience as a valid basis for knowledge, and argues for a unity of knowledges of "hand, heart and brain" [34].

Knowledge and knowledge processes within science are of particular interest for a feminist analysis. A number of questions are relevant to ask around knowledge issues, such as: what knowledge is valid and why? Who can have knowledge? What can we have knowledge about (ontology)? What is the relation between the knower and the known? Who has the preferential right of interpretation and why? And "Whose science? Whose knowledge?" [26]. And finally, not the least: How could it be different? Such questions can throw light on hidden ideas regarding science, and its underlying positions. As we shall see, these questions are important to take into the context of a particular discipline, in this case computer science (see 3 below).

2.4 The situated subject

Most disciplines are based on some, more or less explicitly articulated views of what counts as 'scientific' within that discipline, and thus also on methods, including how to write a research paper. In fact, these traditions, even if seldom made explicit, are very strong and direct the way that for example a graduate student learns to write. Furthermore, these traditions build on epistemological and other theoretical foundations within the discipline, such as what counts as knowledge and how knowledge can be obtained and articulated. Computer science has in many senses adopted much from natural science, regarding how to do science and the role of the researcher etc. This in turn means that it is uncommon to use the pronoun 'I' in research writing. This is a very strong tradition, leaning on the idea of the objective observer, which means that the person doing the work is not even unimportant, he or she should be considered as totally irrelevant for the knowledge produced. Since the author is not present in the text, the knowledge presented is seen as universal by the reader. In other words, the 'knower' is nowhere, all focus is on 'knowledge', thus supporting the idea of disembodied knowledge.

In writing this article, I deliberately break against the rule of the invisible researcher. If I mean to take seriously the concept of situated knowledge, this must also influence my way of writing, of telling about things. The least I can do (and it is not enough) is to make myself visible, as the one who claims to have this knowledge, and who is telling about it. This paper is written from my position, point of view etc. Thus, I am accountable for what is in the text.

3. Feminist theory meets computer science

The body of feminist research in computer science, done by computer scientists, is still small, but it is growing. Interestingly, most of the research is done in northern Europe, not the least within the Nordic countries (some examples are Christina Mörtberg in Sweden, Tone Bratteteig in Norway and Helena Karasti in Finland; see for example [33]).

I see this as developing what I want to call "feminist computer science research". It is important "to raise the profile of feminist research in computing" [4]. Both feminist research and CS are competence areas, but they also bring with them modes of thinking about the world. My belief is that especially feminist epistemological thinking has the potential to enrich computer science. Norwegian informaticians[3] Tone Bratteteig and Guri Verne support this idea. They "see epistemological inquiries to establish alternative understandings of knowledge" as being the most challenging and having the greatest potential for contributing to change in CS [9, p. 60]. I agree with their view:

"*We do not accept the dichotomy between feminism and technology. The challenge is to learn to live with, and possibly harvest from, the contradictions and alleged paradoxes that arise."* [9, p.70]

In this paper, my approach is twofold: I highlight some feminist research that has been done within CS, and that can serve as inspiration as well as basis for future work, and I also point to issues within CS which I believe relevant to research further. I do this by asking questions that I see as important to pursue.

3.1 Knowledge

As I have discussed above, issues concerning knowledge are important within feminist theory. These theories have strong bearings on computer science. The most obvious is in the field of Artificial Intelligence and so called expert systems. Alison Adam, a British AI-researcher has contributed extensively to the critique of AI from a feminist perspective, for example in [2, 3, 4]. She claims that using knowledge and experiences from feminist epistemology, it is possible to get more radical insights into epistemological issues in AI, than when using more traditional approaches [2]. Most critique of AI de-emphasise the cultural production of AI, thus being as Alison Adam sees them, "epistemologically conservative" [4, p. 50].

Traditional criticism of AI concentrates on whether it can create true intelligence, while feminist critique looks to the cultural settings of AI – whose knowledge and what types of knowledge that are represented. What world-view comes with the concretization of knowledge in an expert system? Alison Adam is worried about "the taken for granted nature of the expert and expert knowledge [4, p. 42]. There is, as I discussed above, a difference between propositional knowledge ("knowing that") and skills knowledge ("knowing how"); to represent skills and common sense in AI-systems poses big difficulties. Since different types of knowledge have traditionally been connected to men and women respectively (see 2.3 above), women's traditional knowledge is then denied and withheld.

Alison Adam argues that the epistemology of symbolic AI is based on the Cartesian rationalist view that all knowledge is based on symbolic representations, of which the ideal type of representation is symbolic logic (compare also chapters 9 and 10 in [15]). She recognizes the knowledge in such systems for what it is, that is, a *part* of our knowledge. Rationalist epistemology creates the knower as at once universal and invisible, while feminist epistemology, as I have discussed above, emphasises the standpoint of the observer, and also the role of the body in knowledge production. AI systems reify knowledge. Instead, they ought to reflect plurality of knowers and knowledge.

She sees how two different feminist projects can contribute to change and thus improvements, in the field of AI. The first one is on systems built with traditional AI technology but where the knowing subjects and the limitations of the knowledge are made explicit and where the system is used as an advisor rather than decision maker.

The second, and admittedly more speculative, is within the new field of embedded robotics, which has taken to heart the question of the role of the body in the production of knowledge. However, it does remain to be seen whether the conceptions of embodied knowledge are in fact the same in robotics and feminist epistemology.

Issues concerning knowledge are by no means limited to the area of AI. An equally important question as "whose knowledge is represented in an AI system" is the question "whose knowledge is built into objects in object-oriented design"? Cecile Crutzen and Jack Gerrissen have made a feminist analysis of the

[3]" Informatics is the term for computer science departments in universities in Norway, indicating that the discipline is defined more broadly than in traditional computer science departments." [9, p. 59]

Object Oriented paradigm[4] (OO) [11]. They make a case for making visible what is hidden:

"*OBJECTS should stop acting behind their surface, even if this would render our self-created OBJECTS unpredictable or unreliable"* [11, p. 134).

They claim that object orientation is based on the idea of objectivity and neutrality of representation, as well as the idea that everything and everybody can be represented in terms of objects.

It is interesting to compare this analysis of OO with the views expressed by Sherry Turkle and Seymour Papert ten years earlier [43] where they see OO as potentially revolutionising programming methods and also as challenging traditional ways of thinking and knowing.

Regarding OO I think it is interesting to ask why the word 'object' is used. In fact, in object oriented programming, it is a bit unclear whether objects are acted *upon*, or whether they *act*, in which case it would be more reasonable to call them 'subjects'. Such a change of words, I suggest could entail a change of meaning and conception, which could potentially have effects on practice.

Many other questions regarding knowledge are also important to ask in the context of computer science, such as: what does it mean to "know CS"? How is knowledge created within CS? How are knowledge processes influenced by paradigms[5] and metaphors? How are paradigms constructed and maintained? Can we extend our view of knowledge within CS? Can we cherish epistemological pluralism, i.e different ways of knowing and learning? I see these questions as important for many reasons. For one thing, they relate to the learning of programming, which is one of the fundamentals of CS education. Maria Alsbjer uses feminist theory of knowledge in a study of the processes involved in learning to program [5]. Currently, I am involved in a project with teachers of CS, where we attempt to take this discussion further and put it into the realities of teaching programming at university level [7].

These questions can also relate to the under-representation of women within computing [43, 5, 6]. A re-thinking of CS could mean a broadening of the meaning of "knowing CS", thus potentially accommodating greater diversity in its practices and among its practitioners.

[4] Note that what they criticise is the paradigm of object-orientation at a fairly high level, for example for making analysis of "human worlds", *not* the low level object-oriented programming, used for "realisation of software".

[5] I use the word 'paradigm' here in the loose sense that it is often used within CS, where it is often talked about for example different programming paradigms.

In the sections below, I will attempt to approach these general questions of knowledge through a number of themes, that I find to be of particular interest to focus.

3.2 Paradigms and metaphors in CS

CS is often seen as growing out of and combining other disciplines: mathematics, natural science and engineering. Tensions between these roots exist within the discipline. The three important paradigms identified in [12]: *theory*, *abstraction* and *design*, are in [1] complemented with the concept of *professional practice*. What could this mean for the discipline? These paradigms do in some sense compete with each other, and researching their influence on knowledge production is an important area for future research (see also [8]).

New paradigms or metaphors for computing are surfacing, the most important one today seems to be *interactivity* or *interactionism.* This concept has been discussed by a number of researchers. To take some examples: Lynn Andrea Stein talks about a new computational metaphor: "computation as interaction" [37]; Peter Wegner writes about "why interaction is more powerful than algorithms" [45]; Frances Grundy discusses a new conception of computing that she terms "interactionism" [22] and Heidi Schelhowe sees interaction as a successful approach to development of software [36]. However, these researchers come from different backgrounds, and seem to have developed their concepts in different ways. What could the effects of these emerging paradigms be? In what ways could they support epistemological pluralism, or other ways of knowing? Can different metaphors or paradigms for computation affect the learning of CS?

Another emerging trend, so far mostly within robotics, are concepts of embodied and situated computing (see under Knowledge above). Do these relate to the concepts of embodied and situated knowledge within feminist epistemology?

Paradigms or metaphors of importance within CS will take on a significant role in education. I see the teaching of programming as being of particular importance. What are the paradigms and views of knowledge, CS and programming behind programming courses? Is this visible in the courses or not recognised but taken for granted? What does the currently popular object-oriented paradigm really mean? As it is now, it seems as if the potential power of object-orientation has not brought on significant changes within the teaching of programming, but has rather been incorporated into existing methodologies. If and how a different paradigm or metaphor can promote learning of programming is a question that ought to be of great interest to the whole computer science community.

3.3 Abstractions, formalisations and representations

In computer science, abstractions, formalisations and representations are important. However, there is little discussion about the role of these, and how they are used.

Representations, categorisations and thus simplifications are necessary, but how are they chosen? How is knowledge to be represented within software? What kinds of knowledge should be represented? How can we account for knowing that is situated in social and cultural contexts, so that the situated nature of it does not disappear into universalising and de-contextualising? How do we handle complexities in relation to simplification? What is simplified and how? And as a consequence: whose simplifications, and thus whose knowledge is built into the system? The concept of situated knowledge could be useful in these contexts.

Another important issue for research is the role of abstraction in CS. Abstraction is held to enable methods to be value-free. Computer science focuses on understanding the world via a rationality based in the abstract [38]. However, the products of CS are very concrete. Why is abstract, formal and logical thinking and knowing seen as superior within CS? This question is connected to the issue of how CS relates to mathematics, but I argue that, even though mathematics is important, CS is in many aspects not a mathematical discipline. In contrast, CS could be viewed as concrete science where important aspects are materiality and social practices [10]. Problems can arise when extending abstractions, formalisations and de-contextualisation too far out of their right environments, and applying them in other areas, that do not readily lend themselves to these kinds of descriptions, e g. systems design.

Computer science does require a certain amount of abstract thinking. However, there is no doubt also need (and space!) for what can be called concrete thinking, and not least concrete learning. This could introduce new ideas for gaining knowledge that may make CS more relevant to a more diverse group of people [14]. Knowledge about and acceptance of different types of knowledge construction (see e.g. [5]) is essential if we want to extend our view of knowledge within CS. How would CS gain from truly accepting and cherishing epistemological pluralism [43]?

3.4 Naturalisation

Closely related to representations is the concept of *naturalisation*. In the process of naturalisation, something (an artefact, an idea, a concept etc) is stripped of its origins, context and consequences, and is seen as given, as self-evident.

Susan Leigh Star describes this:

"By naturalization I mean stripping away the contingencies of an object's creation and its situated nature. A naturalized object has lost its aura of anthropological strangeness, and is in a sense "de-situated" in that members have forgotten the local nature of the object's meaning or the actions that go into maintaining and recreating its meaning." [40].

An example of naturalisation within computing is the computer itself. This becomes very clear in meetings with undergraduate students. To most of these, 'computer' does not only mean an artefact, but also a very special artefact – the PC of today! They (and probably most of us) take the construction of the PC for given, not only in the way it appears, but most of all in the von Neumann-model that it builds on, and in the digital technology that is used. The historical contingency of the way that the computer of today is constructed has disappeared. However, there is nothing 'natural' or given with the construction of the present-day computer, not even the digital technology used. For example, Heike Stach [39] shows how von Neumann, in his design of the model, was greatly influenced by the ideas within neurophysiology and psychology (behaviourism) of the time, and not the least of the emerging cybernetics and its ideas of self-regulation and control. He came to formulate his design in terms of the prevailing beliefs of that time concerning how the human brain works. Quite soon, however, the brain came to be thought of in terms of the computer. So – the computer is a brain, and the brain is a computer! (Compare chapters 9, 10 and 14 in [15]).

Using the terminology that Star introduces, the computer is an obvious case of naturalisation, where the choices that were made 60 years ago, and the reasons for these choices, are, if not forgotten, so at least never brought to the fore. What does this naturalisation mean not only for our understanding of the computer, but also for our applications, that are, at the deepest level (machine organisation), completely dependent on this model?

This naturalisation of the computer includes the digital 'nature' of the machine. This too, has in many cases been naturalised, as if it had been the only possibility. Furthermore, the term 'digital' has come to represent much more than merely the digital logic used in computers; for example, it is frequently used to represent virtuality or simply the fact that something is computer-based. In that sense, 'digital' is no longer used as the opposite of 'analogue', and thus that dichotomy has been blurred. This is an example of the effect of a naturalised technological choice reaching far beyond the technology itself, getting adopted and changing meaning in unexpected areas and ways.

What consequences can naturalisation have? For one thing, it is easy to see how everything, from hardware to software tend to be taken as 'natural', as something given, once they have existed for some time. This means that the very special reasons why things are constructed the way they are, are forgotten, and hence there is likely to be a tendency not to question whether this was actually 'the best way' to do something, thus contributing to technical inertia. Designers, machines and software are made invisible, thus hiding the choices that have been made during the process.

This not only affects artefacts, but the making of these as well. Actions and processes are reduced to structures and things, and technology becomes a naturalised object: "In the processes doings and actings are transformed, through collective oblivion, into 'taken-for-granted-ness', which entails that verbs become substantives, with reification as result." [29, p. 147, my translation]. Feminist analysis can contribute to de-naturalisations of the objects created, for example software, in order to understand what intentions and choices that are built into the technology, and can help bring back the active and process nature of technology creation. This will mean that the objects and the processes will become *situated* in the context where they were created, and this situating brings with it valuable knowledge about the different circumstances surrounding the creation.

"Feminist research can become a resource for discerning and revealing the notions that underlie the formation of technology, and in the construction of new notions that facilitate the formulation of alternative technologies, other understandings and ways of thinking that enable a recovery of the poetic rationality that has been lost through the reduction of intentions and choices to things." [29, p. 177]

3.5 Computer science as social and cultural construction

Does CS have what could be called an 'essence'?, does it have an existence 'on it's own'? Or is it socially constructed?

Most feminist researchers acknowledge that both technology and gender are socially constructed. And not only are they constructions that are specific in time and place, they also co-construct each other ([19]) All processes that produce knowledge are situated, socially, culturally and historically, and so is the understanding of gender.

Sandra Harding points out how definitions matter: "*Most engineers, would argue that their technologies are not social at all in any meaningful sense of the term. They have social applications and meanings, but the technologies themselves, i.e hardware* [and software (my addition)]*, are "universally valid" in that they work in any and every culture for what they were intended to do. By excluding from their definition of a "technology" not only its social applications and meanings, but also the knowledge of how to make it, use it and maintain it, they can perpetuate the illusion that technologies are not cultural at all."* [27, pp. 283-284]

This relates to attempts at universalising (e.g. of products of CS), a process that Donna Haraway, along with naturalising sees as connected to the "view from nowhere", the disembodied knowledge. I find it hard to claim universal validity of products of computer science. Software is tightly interwoven with cultural and other pre-understandings of western culture (as mostly interpreted by Microsoft!). This is also discussed by Deborah Johnson (chapter 5 in [15]). According to her, the researcher P. Brey argues that "analysis must be done to 'disclose' and make visible the values at stake in the design and use of computer technology." [15, p. 69], and she briefly mentions how other researchers have shown that the design of search engines is laden with value choices.

In the discussion about computer science and the larger concept of information technology, technology is often seen as an autonomous force, thus implying a technological determinism. However, information technology and computing are not given, they are formed by, as well as forming, people. How systems are constructed depend on who construct them, and what world-view and understandings of knowledge, experience, values and needs they integrate in the development and the final products. Who influences development is thus important to take into consideration [29].

Another important aspect that is often left out is how technology is both created by and creates cultures. The cultures of science cannot be separated from the production of knowledge, these are closely intertwined.

Feminist researchers have studied cultures where science is produced. Sharon Traweek, in a study of high-energy physicists in USA and Japan, has coined the expression of "the culture of no culture" [41], an abstract, depersonalised culture of objectivity. This abstraction excludes everything that belongs to the social world, such as values and ethics, but also the everyday life of meaning. This deleting of the social is also prevalent within computing. In 3.3 above I discussed the role of abstraction within CS. I would suggest that it is possible that the use of (necessary) abstractions in the professional practice can easily lead to abstracting away also ideas, values and meaning, i. e. that abstractions, maybe without being noticed, diffuse into areas where they might not belong, and make us forget and realise complexities and social and cultural circumstances. Hilary Rose [35] sees it as crucial to challenge this 'culture of no culture': "For it is precisely this abstraction, this removal of the human agent from the production process of knowledge, which fosters and informs the lethal culture of contemporary science and technology." [35, p. 26]

Within CS, competing cultures exist. One of these is no doubt the abstract, similar to what Traweek describes, while another (not officially sanctioned but definitely alive) is the subculture of hackers or nerds. However, there are also other cultures, such as those represented by the practitioners of the Scandinavian tradition of systems design[6], participatory design communities etc. These cultures actually challenge boundaries for what is considered to be computer science – but to what degree are these cultures accepted as belonging to CS? And if they are not, this could relate to power, and who has the power to define.

[6] "The central issue of the Scandinavian tradition has been the user involvement in computer based system design. The location where most of the design experiments have done has been working life, in concrete work place settings such as industry and hospitals. The approach has had two trajectories: to participate and influence the democratisation of the working life but also to democratise the design process. […] Today especially in the North American context the approach is called Participatory design (PD)." [13]

4. Feminist inspired comments on *The Blackwell Guide to the Philosophy of Computing and Information*[7]

Feminist theory and the tradition that much of the work in the Guide represents, have different epistemological starting points.

The overall feeling, when reading the Guide, is that most of the chapters (though not all, see further below) assumes a 'traditional' view of "science – as – usual", supporting objectivism, realism and empiricism, and thus firmly founded in the ideas from the scientific revolution. These aspects are held out as foundational for computer science. However, with a different epistemological point of departure, the picture can look very different. My point here will be that of questioning the starting point taken in the Gudie, asking if this view is relevant for computer science, and what the consequences could be. In this context, posing questions such as Whose knowledge? and What knowledge? is very apt, as I will show examples of below. Due to the very limited space here, I have had to restrict myself to only a few comments.

4.1 Information, knowledge and truth

Luciano Floridi, in chapter 4, uses a definition of information where information is considered to have *objective semantic content*. Here, he defines objective as "mind-independent or external, and informee-independent" (p. 42). This means that information exists independently of its encoding and transmission. This view tends to prioritise a view of information as 'object', rather than as process. Primarily, the informee-independence can be interpreted as 'independent of a particular receiver', or assuming a 'standard receiver', in which case the obvious question become: who is this standard observer? However, in the discussion on the General Definition of Information, it is argued that an instance of information "can have a semantics independently of any informee" (p. 45), or, using other words, information does not require an informed subject. Can information really be said to exist if there is no receiver, and if no communication is going on? Is information really just lying around and waiting to be exposed? In a different view, it can be argued that information in itself always involves interpretation (see e.g. the discussion of Heidegger's hermeneutic in

[7] All references in this section refer to pages and chapters in [15].

chapter 25). Furthermore, information requires that data is well-formed and meaningful. But can data be meaningful without an informed subject? Or, is it meaningful for any receiver, independently of who this is?

Fred Adams (chapter 17) argues in the same way as Floridi, that "To be of value to a would-be knower, information must be an objective, mind-independent commodity." (p. 229). Why is it of no value if it is not mind-independent? Adams also concludes that the objective nature of information also means that it is language-independent.

These positions are all about our own standpoints. I rather take the view that information is, to a certain extent, dependent on the position and situation of the informed subject, dependent upon his or hers (situated) interpretation. So I prefer a view that focuses on the subject, and the process, instead of on the object. I agree that data can exist 'in itself', but I hold the position that this is doubtful when it comes to information.

Floridi makes a strong attack on what he terms "decentralised or multicentered approaches" (p. 41), according to which there is no key concept of information. He claims that these philosophers do not accept "the predominance of the factual" (p. 41). And why should they? This is once again a matter of the approach taken, the world-view, the epistemology. I do not think Floridi can show convincingly that a centralised approach is more "right" than a decentralised. He rightly notes that the philosophers he attacks target authoritarian and hierarchical approaches to information (as would feminists do!). The view of information that he argues in favour of has "a core notion with theoretical priority" (p. 41). Using words like 'core notion', and 'priority' certainly suggests a hierarchical and I would say also possibly authoritarian view. In his arguments, he presupposes factual information, i.e. information about 'reality' – but what reality and whose reality does he talk about?

In the introduction to the book "What is the Philosophy of Information", Floridi talks about the 'nature' and the 'essence' of information. Leaning on Donna Haraway, I am highly sceptical to these claims of 'essence'. What would this be? Does information really have an essence? The idea that there is 'essence' in information follows from the view of information as subject-independent above.

Fred Adams (chapter 17), only counts two types of knowledge: empirical and logical-mathematical. This means he seems to agree basically with the empirists stand on epistemology. He claims that:" It is uncontroversial that knowledge requires truth and belief." (p. 228). I am fully aware that the standard definition of knowledge in mainstream (analytical) philosophy is "knowledge = true justifiable belief", but it is *not* uncontroversial when read from a feminist standpoint, mostly because it contains that very tricky little word 'true'! Even if the term is used as meaning something limited and even contingent (Floridi, p. 54, actually uses the term 'contingent truth'), the term is problematic since it brings with it connotations of grand theories and universal, objective truth beyond the subject, something that feminists tend to be very sceptical about. Furthermore, Adams, as well as Floridi, assume the (controversial) standpoint that information must be true, in order to count as information. But truth is a carrier of values, if information is per definition true – then it automatically carries with it the value of accepting something as true.

However, we also have contrasting pictures on knowledge in the Guide. In the chapter on Cybernetics, Roberto Cordeschi points out the merits of the new traditions in AI of embodied robotics: "the intelligence of an agent cannot easily be 'disembodied', since it is also the result of the deep interaction between the agent and its environment" (p. 193). (Compare with the work of Alison Adam, see 3.1 above.)

In the same chapter, Cordeschi discusses the views of what he calls the "new cyberneticans" (p. 194). According to Cordeschi, these researchers see reality as an interactive object, and the *"observer and the observed exist in a perpetually unbroken circular system. These new cyberneticians thus criticize philosophic realism […] These authors consider the activity of knowing not as an act of duplicating or replicating, through internal (symbolic) representations, what is supposed to be already in the outside world, but as a process built up by the observer."* (p. 194).

These "new cyberneticians" thus criticise the idea of gaining knowledge about a reality independent of the observer, a standpoint that I can agree with from a feminist epistemological view. This standpoint gives rise to hermeneutical and constructivist approaches.

However, it seems to me that Cordeschi disagrees with these positions, this is rather clear from the way that he writes somewhat ironically and derogatory of these positions, in their criticism of "the alleged Western 'scientist' …tradition" (p. 195).

4.2 "The computer scientist's world is a world of nothing but abstractions" (Colburn, p. 322)

Taken out of its context, this is a stunning, and quite fearsome statement. However, it becomes clear in the rest of the chapter (Timothy Colburn: Methodology of Computer Science), that Colburn discusses the abstraction in CS of the physical machine, of "the mundane and tedious level of bits and processors" that computer scientists learn to abstract away from. This means some kind of "bottom-up" abstraction, in contrast to the "top-down" abstraction involved in translating real-world problems to be solved into program systems. The kind of abstraction he talks about is of course very important. However, there is the question concerning where, at what level, shall these abstractions meet? At the level of design? Or at the implementation level? Somewhere, a computer scientist must in the end consider the limitations of the machines and system software she/he has at hand.

Furthermore, Colburn, on page 325, argues that "software developers need to become conversant in the analytical tools of philosophers", such as logics, classifications, hierarchies "and other convenient abstractions." From my perspective, that is not what software developers primarily need, instead they need the competencies connected with the domains of use, for example to understand and account for complexity and heterogeneity among users.

Through many chapters of the book runs this thread of computer science as abstract, formal, logical and objective, and its (supposedly strong) connection to mathematics. Is the world understandable and describable in formal terms? My position is that it is not. Whose world is captured in the formal methods/models? As James H Fetzer points out in the chapter on AI (p. 127): "Thinking things and formal systems are not the same. "

So, I believe a fundamental question becomes: what is computing mostly about: formal systems and abstractions or 'thinking things', i.e. people? The answer to this question will depend to a great deal on the view one takes not only of computer science, but of technology on the whole.

To illustrate different views on programming, Colburn (p. 319) takes two quotes regarding programs, one that sees computer programs as mathematical expressions, and another that sees them from the perspective of functionality (the latter quote is from a proponent of the Scandinavian tradition, see 3.5 above). These quotes signal contrasting interpretations as to how computer programs ought to be designed, built, and used. We can ask ourselves which of the above views that is dominating within different computer science communities. My opinion is that the formal view (mathematical expressions) used to be the strongest, but is loosing ground to the more use- and functionality oriented view.

4.3 The supremacy and 'grand narratives' of western science

In the chapter on ontology Barry Smith argues that AI should concentrate on the task of "formalizing the ontological features of the world itself, as this is encountered by adults engaged in the serious business of living." (p. 160). From a feminist viewpoint I react very strongly on such a generalisation. Who are these adults? The AI researchers themselves? To ask questions about whose knowledge and what knowledge is extremely relevant in these contexts. The idea that the experiences of a human being is independent of which human being is selected, is seductive and very dangerous. Feminist scholars have shown that this 'archetype' for a human being is most often a white, western, even middle-class man, and how well does he represent humanity?

If we turn to chapter 24, on Methodology of CS, we see a belief in the success-stories and "grand narratives" of IT. Colburn seems to hold a strong belief in the possibilities of computing methods: "There is no limit to how many real-world processes are amenable to modelling by computer." (p.319). However, we can just as well turn this around, and say, as the critique of AI (not least feminist) has shown, that there is no limit to how many real-world processes that are *not* amenable to modelling by computer.

The chapter on Introduction to Systems Science, by Klaus Mainzer, seems impregnated with the same belief in success stories. This is part of what is sometimes called in feminist theory "the great western enlightenment narratives" which all promise us a good life and some kind of universal solution to all our problems. One example, in the context of society: "By detecting global trends and the order parameters of complex dynamics, we have the chance of implementing favorite tendencies." (p. 38). The question is obvious: whose favourite tendencies?

4.4 Some contrasting views

Charles Hess has written a very interesting chapter on Computer-mediated Communication and Human-Computer Interaction (chapter 6). Here, he among other things discusses the work of Winograd and Flores. They have explored how tacit, non-articulated understandings are built into computer technology. A design of a tool includes certain assumptions, including world-views, and "tools thus embody and embed these assumptions while excluding others." (p. 78). This is what Winograd and Flores express as "in designing tools we are designing ways of being."

(quoted on p. 78). They see much of the world-view underlying design of computer artefacts as 'rationalistic', and instead want to highlight social interaction. Just as Brey (p. 69 in the Guide, see also 3.5 above) and as we have seen, many feminist researchers, Winograd and Flores want to expose and make visible the hidden views and understandings in computer design.

Charles Hess also quotes the feminist researcher Katherine Hayles. She "foregrounds a shift from an objectivist epistemology, based on a dualistic separation of subject-object […] to an epistemology which […] emphasises the inevitable interaction between subject and object." (p. 80). She sees how "embodiment replaces a body seen as a support system for the mind." (Hayles quoted on p. 79).

The most exciting chapter in my view is that by Carl Mitcham on "Philosophy of Information Technology". The views of IT discussed here are very different from those in many of the other chapters. Here, information is much more related to humans and human activities such as language, while in e.g. chapter 4, information is strongly connected to computing and data processing.

Mitcham makes an extensive historical look, and puts information and information technology into its historical context. The ethical and philosophical questions he poses go deep, and far beyond the surface, but into the very creation and sustaining of information technology. He provides an in-depth interpretation of Martin Heidegger on IT. This view emphasises the processes of interpretation of information, and that a more holistic perspective is necessary: "all information technology is part of a larger life-world and cannot be understood apart from such an implicit whole." (p. 333). Heidegger claims that information technology not only reveals, at the same time it conceals. This thinking casts another light on the use of formalisms and de-contextualising, and the obvious question becomes: what is concealed and hidden from view?

Following Heidegger, Jaques Derrida (whom Floridi attacks in chapter 4 on information, see 4.1 above), proposes to bring light on hidden assumptions, to "that on which they depend without knowing or acknowledging it" (p. 335) in information technology design.

To summarise, we see how very different views on computing can be seen depending on the position of the author. In this Guide, the chapters that are closer to a 'pure' computer science tend to lean towards a preference for the abstract, logical and formal, while the chapters that discuss the broader aspects of information technology or use of computers talk about embodiment, interaction, interpretation and hermeneutics. We can see how the views in these different chapters reflect different philosophical traditions, and thus different epistemologies. As is obvious from my comments, the feminist epistemologies I start in, have much more in common with this latter tradition, even if feminists also criticise some of the points that are made in this philosophical tradition.

5. Concluding remarks

This article is full of questions, and has very few, if any, answers. One important point for feminist research is to ask questions, thereby promoting reflexivity on behalf of the practicing scientist. Asking questions is a way of starting a reflective process as well as a way to communicate.

These questions have implications for practice, such as what we convey to students in programming classes. What (implicit) assumptions and commonly accepted views underlie the knowledge processes in CS, e. g. (teaching of) programming? As for curricula and syllabi, what assumptions about knowledge and the subject do they presuppose? I believe that reflection around issues of knowledge is important for every discipline, especially for teaching and for meeting potentially new groups of students. Can we, as computer scientists, by becoming aware of our own views of knowledge and understanding, also become aware of, respect and accommodate for, greater diversity among students, and their backgrounds, interests, motives and understandings? Can we thus, in the long run, change our discipline into one that is more attractive to a broader range of students, for example women? I believe that this can be one contribution to the large task outlined by Maria Klawe which I deem to be of great importance for computer science to pursue [28, p. 68]:

"We need non-nerds in computer science, so let's figure out the proper approaches to integrate their talents and perspectives into our field."

I believe that one of the most important things for feminist research in technology in general as well as within computing, is to work at the broadening of the concepts and understandings of technology, for example to include use and social practices. There is nothing inevitable about how computing is constructed, thus it can be re-visioned and re-conceptualised. Feminism is a resource that can be used to formulate alternative goals, visions and dreams about our existence, feminist research may contribute to re-configure, re-formulate or start to give technoscience other directions [32].

6. References

[1] ACM and IEEE-CS, 2001: *Computing Curricula 2001, Computer Science Volume*. http://www.acm.org/sigcse/cc2001/ [2004-05-01]

[2] Adam, A, 1994: "Who Knows How? Who Knows That? Feminist Epistemology and Artificial Intelligence." In Adam A. et al (Eds) *Women, Work, and Comp-uterization: Breaking old boundaries, building newforms*. Amsterdam; NewYork, Elsevien, 1994, pp. 143-156.

[3] Adam, A, 1995 "Artificial Intelligence and Women's Knowledge: What Can Feminist Epistemologies Tell Us?" In *Women's Studies International Forum*, vol 18, no 4, 1995, pp. 407-415.

[4] Adam, A, 1998: *Artificial Knowing: Gender and the Thinking Machine*. London, New York: Routledge.

[5] Alsbjer, M, 2001: *Att hitta ingångar i formandet av programmeringskunskap. En diskussion kring möjligheter och svårigheter i grundläggande programmerings-undervisning inom högskolan*. B Sc thesis, Blekinge Institute of Technology.

[6] Björkman, C, 2002: *Challenging Canon: the Gender Question in Computer Science,* Licentiate thesis 04/02, Blekinge Institute of Technology.

[7] Björkman, C, Dackman, C, 2004: "Knowledge and Learning in Computer Science from a Gender Research Perspective" – a project with Computer Science teachers" Paper presented at *Symposium Gender and ICT: Strategies of inclusion*, Brussels January 20.

[8] Björkman, C, Trojer, L, 2002: "Computer Science and its Paradigmatic Basis –Broadening Understandings through Gender Research from Within" in [6].

[9] Bratteteig, T, Verne, G, 1997: "Feminist or merely critical?" In Moser, Aas (eds) *Technology and Democracy: Gender, Technology and Politics in transition?* Centre for Technology and Culture, University of Oslo.

[10] Clegg, S, 2001: "Theorising the Machine: gender, education and computing" In *Gender and Education, vol 13, no 3*, pp. 307-324.

[11] Crutzen, C, Gerrissen, J, 2000: "Doubting the OBJECT World". In *Charting a course to the future*. Proc. of Women, Work and Computerization Conference, Vancouver, Canada.

[12] Denning, P, et al, 1989: "Computing as a Discipline". In *Comm. of the ACM*, vol 32 no 1 pp.9-23.

[13] Elovaara, P, 2004 (forthcoming): *Angels in Unstable Sociomaterial Relations : Stories of Information Technology.* Ph.D. thesis, Blekinge Institute of Technology.

[14] Estrin, T, 1996: "Women's studies and computer science: their intersection". In *IEEE Annals of the history of computing,* vol 18 no 3, pp. 43-46.

[15] Floridi, L (ed), 2004: The Blackwell Guide to the Philosophy of Computing and Information. Malden, Oxford, Carlton: Blackwell Publishing.

[16] FoxKeller, E, 1992: *Secrets of Life, Secrets of Death*, Routledge, London.

[17] Fox Keller, E, 1995: "The Origin, History and Politics of the Subject Called "Gender and Science", in Jasanoff et al (eds), *Handbook of Science and Technology Studies*, Thousand Oaks, London, New Delhi, Sage Publications, pp. 80-94

[18] Fox Keller, E 1995: *Refiguring Life – metaphores of twentieth-century biology,* Columbia Univ. Press.

[19] Gansmo Jøsok, H, 2002: "Samfunnsproblemet 'jenter og data'". In *Kvinneforskning* 1/2002, pp. 10-25

[20] Grasswick, H, Andersen, N*: Bibliography Feminist Epistemology*. http://ucsu.colorado.edu/~brindell/soc-epistemology/Bibliographies/Feminist_Soc_Epis/index.htm [2004-05-01]

[21] Grint, K, Gill, R, 1995: *The Gender-Technology Relation: Contemporary Theory and Research,* London: Taylor & Francis

[22] Grundy, F, 2001: "A New Conception of Computing: Interactionism Replaces Objectivism". Presented at *GASAT 10 Conference*, Copenhagen, July 2001 http://www.keele.ac.uk/depts/cs/staff/a.f.grundy/home/interact.htm [2004-05-01]

[23] Haraway, D, 1991: *Simians, Cyborgs and Women:The Reinvention of Nature*, Free Association Books.

[24] Harding, S, 1986: *The Science Question in Feminism*, Cornell University Press

[25] Harding, Sandra, 1987: *Feminism and Methodology*, Indiana University Press.

[26] Harding, S, 1991: *Whose Science? Whose Knowledge?* Cornell University Press.

[27] Harding, S, 1996: "Multicultural and global feminist philosophies of science: resources and challenges." In Nelson, L.H & Nelson, J (eds) *Feminism, Science and the Philosophy of science*. Great Britain: Kluwer Academic Publishers, pp. 263-287.

[28] Klawe, M, 2001: “Refreshing the nerds”. In *Communications of the ACM,* vol 44, no 7, p. 67-68.

[29] Mörtberg, C, 1997, “*Det beror på att man är kvinna...", Gränsvandrerskor formas och formar informationsteknologi*. Ph.D dissertation 1997:12, Luleå University of Technology.

[30] Mörtberg, C, 1999: ” Technoscientific challenges in feminism”. In *NORA (Nordic Journal of Women's Studies)* Volume 7, Issue 1.

[31] Mörtberg, C, 2000, ”Teknikvetenskap och genusforskning eller Creating Better Worlds Through Imaginations” in Trojer (ed) *Genusforskningens Relevans*, *Slutrapport från integreringsarbete i åtta svenska forskningsråd,forskningsrådens expertgrupp för genusforskningens integrering*. Sthlm.

[32] Mörtberg, C, 2003: “In Dreams Begins Responsibility – Feminist Alternatives to Technoscience” In Mörtberg, C & Elovaara, P & Lundgren, A (Eds.), *How do we make a difference?,* Luleå: Luleå University of Technology. pp.57-69.

[33] Mörtberg, C & Elovaara, P & Lundgren, A (Eds.), 2003: *How do we make a difference.*. Luleå: Luleå University of Technology.

[34] Rose, H, 1983: “Hand, brain and heart: Towards a Feminist Epistemology for the Sciences”, *Signs, 9*, 1983, pp. 73-90.

[35] Rose, H, 1994: “Feminist Standpoints on Science and Technology”. In Gunnarsson, E & Trojer, L (Eds), *Feminist Voices on Gender, Science and Ethics*. Luleå: Centre for Women’s Studies, Luleå University of Technology. pp. 15-30.

[36] Schelhowe, H, 2004: “Computing Science and Software Development: Paradigms of Mathematics, Engineering, Interaction” Paper presented at *Symposium Gender and ICT: Strategies of inclusion*, Brussels.

[37] Stein, L A, 1999: “Challenging the computational metaphor: implications for how we think” In *Cybernetics and Systems*, vol 30 pp. 473-507.

[38] Stepulevage, L, Plumeridge, S, 1998: “Women taking positions within CS” In *Gender and Education vol 10 no 3* pp. 313-326.

[39] Stach, H, 1997: “The construction of the von Neumann concept as constituent for technical and organic computers”. In *Women, Work and Computerization. Spinning a Web from Past to Future. Proceedings of the 6th International IFIP-Conference*, pp 423-33.

[40] Star, S L, 1994: “Misplaced Concretism and Concrete Situations: Feminism, Method and Inormation Technology” In *Gendet-Nature-Culture Feminist Research Nerwork, Working paper 11*, Odense University.

[41] Traweek, S, 1998: *Beamtimes and Lifetimes: The World of High Energy Physicists*, Harvard University Press.

[42] Trojer, L, 2002: *Genusforskning inom teknikvetenskap - en drivbänk för forskningsförändring* [Gender research within Technoscience], Stockholm: Högskoleverket.

[43] Turkle, S, Papert, S, 1990: “Epistemological Pluralism: Styles and Voices Within the Computer Culture”. In Signs: Journal of Women in Culture and Society, 16(11).

[44] Vehviläinen, M, 2000: “Gender and information technology”. In Mörtberg (ed): *Where do we go from here? Feminist challenges of Information Technology*. Luleå University of Technology.

[45] Wegner, P, 1997: “Why Interaction is more Powerful than Algorithms. In *Communications of the ACM,* vol 40, no 5, pp. 80-91

Attitudes and Research Methods in Information Design

Carina Andersson, Bengt Olsson and Rune Pettersson[1]
Department of Innovation, Design and Product Development
Mälardalen University in Eskilstuna, Sweden
carina.andersson@mdh.se

Abstract

This article presents a discussion and reasoning about how a design process, within information design, and a scientific process can co-operate. The article is based on unstructured interviews. The interviews are conducted with five professional respondents skilled in the handcraft of information design. The article also presents an analysis of scientific publications and a study among the research team within Information Design at Mälardalens högskola in Eskilstuna, Sweden.

The study is a base for a discussion and reasoning about how a design process and a scientific process can co-operate within Information Design.

The craftship and the theories ought to enrich each other. An information designer has to be aware about her/his activities and the knowledge that can be brought forth with a specific intention. To what extend Information Design is science can depends on our definition of science, the scientific ideal and the theories we use as well as our basic view of how to build knowledge.

1. Introduction

Information Design is a multidisciplinary field of study at Mälardalens University College in Eskilstuna, Sweden. Information Design is concerned with how to make information sets understandable to humans, for instance as manuals, exhibitions, web-sites and fire alarms. An educated student in Information Design will master how to plan, analyse, prepare, shape and evaluate an information set. (Pettersson. 2002, Jacobsson. 2000, IIID. 2004-02-13, Horn. 2000).

A student who is studying Information Design requires craftsmanlike skills with a theoretical scientific attitude. Students within Information Design have to face up to that one person, an information designer, should attain competence in both handcraft and scientific methods. Is there a paradox?

During our discussion concerning science within Information Design we raised questions about design methods, design theory, design science etc. Does design belong to nature science or humanities and social science, should design be studied with a hermeneutic attitude or a positivistic? In design research we have come across statements like that “no designer are working after one method”, or “design is nothing in itself but part of other sciences.” Other statements are that “it is impossible to create fundamental laws within design because design is about creativity” or that “design is a practice.” (Snodgrass and Coyne. 1997, Cross. 2002, Willem. 1996, Nijhuis & Boersema. 1999).

Since Information Design is a multidisciplinary field of study, research within Information Design is a part of other sciences as the science of cognition or the science of communication. Within our university education it happens that the scientific discussion only concern the framework and the final set of information. Target-group analysises, usability test and evaluations are done to study if the information design fullfill its purpose. Knowledge about how to deal with a scientific attitude within information design is a necessity at our department. Therefore we are reasoning about what kind of methods can be used in Information Design research and what kind of scientific attitude can be suitable and what are the suitable methods within Information Design?

Snodgrass and Coyne (1997) claim that: ”*Designers are truly designing when they are so absorbed in the task that they are not aware that they are designing.”* (Ibid). If a designer is totally consumed up by his creativity and not aware about scientific methods during the design process, then it could be impossible to educate students into good craftsmen with a conscious scientific attitude. Therefore we

---

[1] Bengt Olsson and Rune Pettersson are discussion partners. They have also contributed with: 1.1., and 4.2., and parts of 2. parts of 2.1 and parts of 4.1.

discuss in what way the design process, creativity and scientific process of research can co-operate.

1.1 Definitionof Information Design

In this article the term Information Design is defined as (1) a multidisciplinary topic, which comprises a holistic view of techniques and processes adapted when shaping a set of information as well as studying the usage of a set of information. The term Information Design also refers to a group design topics, a group of related subjects. The purpose of those subjects is to help humans to understand how something works. For humans in need of information, well-performed information design makes everyday life easier.

Information Design, as a multidisciplinary topic, shares points of view with other areas of knowledge. The most important ones are language, art, aesthetic, communication, information, cognition, economy, law and media and technique. (Pettersson. 2002). Information Design is also an integrator, influencing topics like Product Development, where well-performed information design can be significant for how actors in the product development process receive, interpret and perceive a project as well as project documentation. (Andersson & Elfving. 2003).

2. Design

The concept of design got a widen content in the end of the 20-century. Focus has now changed from traditional values of aesthetic and the art industry to functional values in modern industrial mass production and concepts of usability. In the modern industry the correlation between economy, material, human, environment, organisation and technique becomes even more important factor than before. Design is not only about physical product, design also contains the shape of a message, physical environments, communication, meetings, technical systems and virtual environments.

In this article the concept design is (1) processes used when shaping products, environments and systems. The shape should support the usability, the shape should be aesthetic, and it should be adapted to economical, material, human, environmental, organisational and technical demands. The term design also includes the result of (1).

The concept designer, in this article, is a person who leads a project, visualising products, environments or system, both in mind and in performance. Cross (1996) declares that design contains some of the higher cognitive abilities as creativity, problem solving and it unites parts to a whole.

2.1. Creativity

It is rather complicated to describe creativity as a process, because it is not linear sequences of steps or phases. This results in complications when people compare creativity with processes for instance a design process. The complex of problems can be in the confusion of descriptions of the word creativity. Hagoort, G. (2000) refers to G. Wallas, who developed the well-known PIIV-model, which describes creativity as concisting of following stages: Preparation, Incubation, Illumination and Verification. But there are other examples of descriptions. One is a pragmatic perspective:

A female dancer describes her creativity as:

A) The starting point can be a word, an image or a piece of music.
B) The brooding deeply thinking phase. The limitations of the area starts to appear.
C) The concentration phase. The Idea gets a concrete shape.
D) The risk phase. The performance /piece is presented for an audience.
E) Finally it shows if it has or if it transfer the intended value or quality.

A composer describes his creativity as:

A) Inspiration – an unscientific idea with tones and rhythms – a music-idea. The inspiration gives impulses to creation.
B) The impregnation – depending on the style of the first phase, the music takes form. This form/shape is not an expression of existing standards but a new language!
C) Enforcement/execution, the composition is presented. When the piece of music is presented the composer do no longer see his music as his own. The purpose was to create new music, not repeat an old one (Ibid).

You plan to be creative, but you do not plan the creativity. To use ready-made methods when sovling a problem is not creative. Neither is the usage of a tool, but the tool can be used to express creativity (Raudsepp. E. (1993). The information design tools on the disposal are language, images and shapes. We can say that the creativity is within the designer's activity and knowledge of how to combine language, images and shape.

2.2. A Design Process

Design processes apply to processes when working with design. Those processes describe the planning and shaping of a product, for instance an information set. A design process is multi-faceted and one way to describe it is as a communication process. A general description can be:

analysis of/defining problems
analysing requirement,
planning the work/project,
synopsis,
development,
production,
evaluation,
making decision,
the finale product/commission

During a design process there are several phases when decisions have to be made, as well as a recurring of tests and reviws, this to aviod eventual defects of the product. An evaluation is a way of examin the design, the goal is the final product, the assigner can be a customer or a company, and the target group is the users. (Lundeqvist. J. 1995, Nijhuis. W & Boersema. T. 1999, Pettersson .2002).

Design activity and design processes are not the same. To design can be described as the sum of the designer activities and those activities results in ”*realizing the transformation of designing in the design process /.../ also computers should be considered as transforming systems*.” (Hubka & Eder. 1997). A design process also includes production, for preparing for press, a phase when the designer is no longer active. The designer’s activity is just one part of the design process.

2.3. A Research Process

There are also several ways of describing a research process. A simplified picture of a research process can be:

problem/complex of problems,
questions/hypothesises,
literature study/facts, analysises,
theorizing/methods (induction and/or deduction),
testing of for instance hypothesises,
analysis, evaluation and discussion of the result

Recurring of reflections are done between the diverse phases. Evalutation of the result leads to future research and the goal is to give birth to new knowledge (Roozenburg & Eekels. 1995).

In research within Information Design we work with every specific research question, or problem. From a description of complexity of problems to a final product the research process can be decribed in seven phases: (1) analysis, (2) project planning, (3) litterature studies, (4) collecting of data, (5) processing of data, (6) discussion and (7) publishing (Pettersson. 2002).

2.4. A Design Process and A Research Process

Nijhuis and Boersema (1999) bring up the issue that a graphical designer and behaviour scientist can co-operate and produce a successful result. Sometimes conflicts are raised because of professional differences in how to deal with time and money. Even though there are several similarities between a graphical designer and a behaviour scientist. Many of the similarities can be found in the processes, the design process and the process of research. *”Not only is it possible to construct congruent strategic models of the two disciplines, but it is also shown that the corresponding tactics are remarkably similar. Differences exist only at the operational level, where specific skills and methods are used to achieve intermediate results*.” (Ibid. s. 21-41).

3. Design and Science

The meaning of concepts concering design science is debated. Cross (2001) describes three different ways of approaching design and science. One should distinguish between scientific design, design science and science of design.

Scientific design has its foundation in industrial design, and deals with methods, intuitive and non-intuitive, in modern design practice. Science of design is the scientific analysis of the design activities preformed via scientific methods. Cross (Ibid) defines that design science deal with the organisational, the rational and wholly systematic approach to design. Design science does not only bring forward the scientific knowledge of artifacts but also design as a scientific activity in itself.

The critisism of design science is based up on the caim, for instance, that no one designer is following one and only method, or that there are no fundamental laws within design, scince design is a practice (Snodgrass and Coyne. 1997, Cross. 2002, Willem. 1996, Nijhuis & Boersema. 1999).

3.1. What is a Theory?

A theory is a number of assumptions or statements that conceptualises diverse phenomen, and systematises our knowledge about them (Nationalencyklopedin. 040319). According to Hooker (1991) is a theory not only a conceptual framework, an enclosed area. Carl von Linnés system built for categorising arts, is not a theory, subsequensly it does not state anything. A theory illustrates how things/something is and why.

A theory is as a tool for the human mind, just as equipment can be a tool for an action (Lundequist. J. 1995).

3.2. Design Theory

Design theory can be viewed in two perspectives, a "dogmatic design theory" and a "zetetic design theory". The dogmatic theory deals with how a model is transferred to a concrete object. A zetetisk perspective relates to the understanding of the performance, in other words, how the model is created. The actions will be perceived as a way of communicating meaning (Ramirez. 1995).

Design theory can be defined and handled differently depending on how we reflect upon design. Let us focus on how design can be perceived, then design theory will deal with socio-phycological phenomenon. It is for instance the political and cultural context, which influences how humans understand and create meaning. The interpretation of design is within the knowledge of the signs that are surrounding us (Rampell. 2002. s. 14. ff). If we approach design as a craftswork, design theory will deal with how to organise the knowledge of design as a practise (Hooker. 1991). If we reflect on design as Le Corbusier, a way of thinking (Nylander. 1999), design theory could deal with the actual "thinking process" when creating a design.

There are doubts though if the design process can be theorised about since *"...one cannot theoretically organize his knowledge of how to design."* (Hooker. 1991). How a person is designing is supported by several theories, since design is a part of different supporting sciences (Ibid).

If theorising about design/design process is possible depends on how humans look on knowledge. Ramirez (1995) claims that design theory represents a reversal of theory of knowledge. If theory of knowledge is about how reality can be perceived and how our ideas meet the world, a design theory is about how reality is created and how our ideas and experiences can shape/change the world.

3.3. Design Science

Some scientists believe that a design science is impossible and some scientists think that it has a place in the world of research. You may wonder: "What is design science?". For instance in Sweden, design science is a topic at "Mitthögskolan and at department "Lunds tekniska högskola". Design as a science was given expression to during the period of functionalism, around 1920, when the design association, De Stijl, argued that creating new products required methods, obective systems (Cross. 2002). During this period of time architecture was systematised and organised by Le Corbusier.

Herbert Simon introduced design science as a concept in 1969, in the textbook "The design of artefacts". The introduction of design science was a way of separating the research in academic design topics from the practical artistic activity, used when developing utility goods (Margolin, 2002, s.235). Design science was movement of forming a doctrine about the design process (Cross. 2000), and a doctrine of "the logic of artifacts which is related to /…/ the logic of the natural world." (Dasgupta. 2003).

Dasgupta (Ibid) presents Simon's work. Simon discussed principles and laws, similar to the concept of usability. Some of those principles and laws were: "*satisficing, heuristics, search and bounded rationality and satisfactory*" (Ibid). In design solutions there are grades of satisfaction, both economically and during the actual usage. Something that seems satisfying in the beginning might not be to satisfaction in the final result. "*Thus, designs, in general, constitute satisficing solutions to design-problems*" (Ibid). Notice that Simon was active at the time when the concept of desing increased its content from traditional aestetical values to functional values in modern industrial mass production and concepts of usability.

There are scientists, who connect design science to the positivistic formal logic, as design scientists try to establish an underlying philosophy concerning the meaning of design[2].

4. Science and Scientific Attitude

Science is often directly related to scientifical method. To explain the reality and predict eventual insidents is one of the main tasks (especially within natural science). Compared to the humanities that deal with the understanding of actions and behaviour of humans (Nationalencyklopedin. 040301, Lundequist. 1995, Dodig-Crnkovic. 2003).

What can be considered as science? Let us examine the newly established field of "educational science" (utbildningsvetenskap). There are different opinions about the being of educational science. One opinion is that education science comprises the science and the research attached to teacher training (Högskoleverket. 040311). It is possible that educational science is mainly a political creation, something we can relate to some of the social constructivists, who argue that building of knowledge from geography to mathematics is a human invention (Nationalencyklopedin. 040301).

Research is the work of developing theories, methods and concepts to broaden the existing knowledge and give birth to new knowledge. OECD (Organisation for Economic Co-operation and Development) has accepted that research can be performed as:

- fundamental research (seeking of new knowledge and new ideas),
- applied research (aiming at that the application), and
- developing work (which systematically and methodically use scientific knowledge to create new/improve products/processes) (Ibid).

Within the field of research there are two main traditions, the positivistic tradidion and the hermeneutic. These traditions represent different scientific attitudes and ideals.

2 "*...Underlying their desires the hope that such a universal design philosophy will establish a consistent unifying foundation for design in the same way that the logical positivist approach to science provided a foundation for its many different branches*." (Kent MCPee. s. 16).

4.1. A Positivistic Scientific Ideal

The positivists declare that the human knowledge is limited to the experience of the senses, and to the general assumptions based on that experience. The positivistic main purpose is to contribute to universalise science and the human knowledge. It means to observe, measure, describe, systematise, correlate in a taxonomy/ontology and be able to predict and reproduce behaviour.

The scientific manifest of the logical positivists can be summarised into:

1. Integration of the scientific knowledge, which includes the idea of a universal language.
2. Criticism of metaphysical statements and delimitation of cognitive statements (separating the meaningful cognitive statements from the meaningless). This includes the principle of verification and the distinction between synthetic and analytical statements.

The applied method is the logical analysis. This method was included in the formal logic ideas during 1900. Actors of formal logic claimed that sensory impressions do not need to be interpreted. The sensory impressions are neutral observations. These observations are used as foundation to all knowledge.

The logical analysis was considered as the one and only philosophical method. The method can be used: (A) *negatively*, to eliminate metaphysical statements from "empirical sciences", science built up on experiences. (B) *positively*, to elucidate scientific concepts and methods for declaring how the human knowledge is based on data hold by experiences (Johannessen, K. 1999). It is important to understand that positivism is a scientific attitude and an ideal. You can be a mathematician with a hermeneutic attitude, for instance in the doctrine about how humans teach and learn mathematics (within mathematics and didactic).

4.2. A Hermeneutic Scientific Ideal

Due to the fact that there are several parallel concepts in the hermeneutic tradition, it is not as easy to describe as the positivistic tradition. In the following we present the growth of hermeneutic tradition from three perspectives.

According to Wilhelm Dithey a human being can perceive phenomenon in physical, mental or spiritual way, because a human being can develop concepts or symbols. We do not need to have the object or the phenomenon physically present. A concept or a symbol has a representative function in our consciousness. This is called “The dialectic hermeneutics”.

According to Karl-Otto Apel it is more fundamental to understand something, than it is to explain the cause of it. In order to explain the cause of a phenomenon, we first have to understand what it is. We transform from concept to comprehension. How humans understand is a historic and a linguistic retention by the traditions of the culture. Interpretation and appraisal influence understanding. A concept can be understood when humans can watch and listen to how it is used. The pragmatic dimension in this process connects understanding to the users way of acting and ”form of life”. This is called “Transcendental-pragmatic hermeneutic”.

According to George Henrik von Wright all phenomena that constitute the human culture are characterised by meaning and intention. The purpose of a specific action cannot be explained if we fail to understand the situation in which the action is taking place. In this hermeneutic tradition the understanding of the situation is crucial. By the way of conscious actions, meaning can be obtained and learned. This is the pragmatic view of meaning. This is called “Critical-analytic hermeneutics” (Johannessen, K. 1999).

4.3. Design and Hermeneutic

Niiniluoto (2001) discuss design science with a hermeneutic attitude. Design science is not about describing how things are or to prevent eventual incidents. Its about explaining how something should be to reach a certain purpose or goal. This give birth to knowledge in how planned actions can reach to a certain purpose. This socalled ”mean – ends relation” is based upon the criticalanalytical hermeneutic

How to handle information design as a practice and design as conscious choices? According to Von Wright (1971) it is impossible to examine a person’s intention without examining how those intentions transformed into an action. Von Wright’s intention model deals with action and intention.

1) *The individual (I) have the intention to reach a goal (G).*
2) *According to I´s opinion it is necessary to make a specific action (A) in order to reach G.*

---

3) *When the right situation is present then I start doing A* (*Ibid*).

A conversion of this model can be described as: *If you (I) want G, and believe that you are in situation B, you ought to do X.*” (Niiniluoto. 2001). X is a recommended action to reach a certain goal. Niiniluoto argues that the describing sciences do not accept such conclusions, since it is an expression of how things should be, ”*but it may be a part of what we may call design science*.” (Ibid. s. 375).

Snodgrass and Coyne (1997) argue that design work is hermeneutic and follows the hermeneutic spiral, since design is based on the relation between the whole, the parts, and interpretation in every step in the process. Design can not be performed in a positivistic attitude and with a formal logical alignment. Design is not based upon problem solving, but on interpretation and the comprehension of a situation. Snodgrass and Coyne claims that design is adapted to a certain situation, where object and subject is inseparable, just as in the philosophy of human sciences. Design is a symbiosis of questions and answers, a so-called *question-answer structure*.

Hermeneutics discuss that there are at least three ways of understanding a phenomenon: bodily, mentally and spiritually. Data, within humanistic and social sciences, should not compare with nature science. It is not enough with observation. Hermeneutics claim that a scientist has to understand the data. It can be compared to research within Information Design, since information can not be achieved until a user received, interpreted and understood the data.

5 Information Design

Data can be defined as values, processes and facts. Information can be a process, knowledge or an artifact/set (Buckland. M. K. 1991).

In Information Design Data is defined as values. Data 3 can for instance be a number (the age of a child), temperature, or metres. Data 3 is a value until it is organised and embodied in an information set. The information set is a representation of the data, how this representation of data can be interpreted and result in meaing is a process in human’s mind. Hence information, in Information Design, is the result of humans' interpretation and attempt to produce meaning.

5.1. A Theory

If humans should be able to generate meaning it requires consciousness or unconsciousness knowledge of the codes that are shaping the world (Rampell. 2002). Humans are meaning-creators. A receiver creates meaning in interplay with, for instance the surrounding culture, and political context. We are looking for patterns to interpret a content. The interpretation is affected by our memory, per understanding, context etc. (Arai, 2001, Goldstein. 1998, Rampell, 2002)

It is not only the receiver who interprets. The sender (in this case an information designer) is interpreting while selecting, structuring and organise information. There is always two interpretations, the sender's and the receiver's (Floridi. 2002.).

A theory in information design can discuss how we can understand humans' interpretation, building of meaning and knowledge.

5.1. Research

Most of the research in Information Design is applied research. Information Design is a multidisciplinary field of research. We use critically selected research results from other topics, to develope directions for shaping an information set that fulfil demands of communicativity and economy. Experiences in Information Design are applied practically within areas interested in well-done information sets. Evaluations give feedback when receivers tell us if the information set was succsessful or not. The evaluation can be used in Information Design as a topic and in topics as communication science and cognition science. This is how Information Design continuous contribute to a cycle of information, which leads to new knowledge. (Pettersson. 2003).

Within nature science, laws can be categorised as fundamental, empirical and observable phenomenons. Some scientist's claims that it is impossible to create fundamental laws in design, since design is innovative and creative. There are fundamental laws, principles and guidelines within Information Design. A fundamental law is that all documentation should be current, understandable, correct, and relevant for a target-group. A principle is the RLR-principle. An information set should be readable, legible and have reading value. Guidelines can guide us when choosing suitable colours for overhead- and PowerPoint-presentations (Pettersson. 2002).

6. Method Used for Discussion

During six gatherings we have discussed and reasoned about research methods and attitudes within Information Design. The discussions lead to a mapping of design education in Sweden, a categorising of design topics, a literature study within design theory, design methods and design science. The discussions also lead to an investigation at our own department[3], and an analysis of two scientific magazines. In this article the literature study, the investigation and the analysis are presented. We use these methods to obtain a greater understanding and create a foundation for the discussion.

How concepts are described in theory and how human uses them in practice can vary (Johannessen, K. 1999). To investigate tacit knowledge at our department at to find out how concepts are practised; five 1,5 hours long interviews have been conducted with teachers, as competent craftsmen. The questions were open and aimed, and based on the concepts: scientific attitude, methods, design process and creativity. The interviews were followed-up by a group discussion among the staff active in Information Design.

Since the topic Information Design concerns theory based on practically performances and scientific founded knowledge, this article includes a study among the research team in information design. The questions, answered by the teachers, were answered via e-mail by the research team. IVLA Selected Readings and ID-Journal[4], have been analysed to find answers of what kinds of methods are used in Information Design research.

7. Results

Herein after is a summary of the results and the conclusions of the two scientific journals and the research team. At the next page you can read about the conclusions of the interviews.

There are different methods presented in **IVLA Selected readings**. The 30 investigated articles presents purposes that are descriptive (describes the reality via for instance observation), constructive

[3] The deparmtent of Innovation, Design and Product Development at Mälardalen University College, in Eskilstuna Sweden..

[4] The International Visual Literacy Association and Information Design Journal

(developing of products), normative, prescriptive but not predictable. The methods are often qualitative and conducted as, experiments, case studies, studies of historical information, theory-building, mappings, usability testing, project descriptions, presentations of questionaiers, personalisty test in combination with observation. The articles consider how humans apprehend and behave.

The study of eight articles in **ID-Journal** shows a multitude of research areas and methods. The purposes are for instance normative and constructive. The methods are for instance experiments and simulations, questionnaires, quantitative and qualitative analysis and philosophical concept inquiry. The methods where mostly quantitative in this specific ID-Journal.

In all the articles a qualitative study seems more common. The descriptions of methods and concepts differs, the methods are not unison and not even comparable. It can depend on that the articles are conducted in different research areas as engineer science and educational science. Within these areas diverse concepts can have similar meaning and vice versa.

The conclusion is that in the two scientific magazines a multitude of diverse methods can be used as research methods in Information Design.

The research team uses, for instance, methods like experiments, building and testing of prototypes, questionnaires, interviews, and observations. The methods are both quantitative and qualitative, with a qualitative majority. The descriptions of methods and concepts differ. It can depend on that research is conducted with different approaches, for instance, behaviourism or action research. We know that employees in the research team origin from different scientific traditions. The similarity is that all employees in the research team are working with how humans apprehend and behave.

The conclusion is that the research team in Information Design work with a multitude of methods from diverse scientific traditions.

Questions have been asked when employees in the research team are satisfied with their research result, and how they estimate quality in their field of work. Joint statement is that they are satisfied when their research results are reviewed and tribute by expertise. As one researcher describes: “The easiest way to say it is that an article that has been published in a qualified scientific magazine has good quality.” One researcher says that he is never totally satisfied since every new result raise new questions. “There is always more to find: One is never ready.” One PhD-student describes good quality as building knowledge and know-how within a certain area. Another PhD-student writes that good quality is a durable development, “A research result always has to be developed. I am satisfied with the result when I can see that there is a connection to other research.”

Employees in the research team link satisfaction of the results with good quality. They think that a good result and good quality is for instance, a verifying or falsifying of a hypothesis and if their work can lead to new knowledge.

7.1. Conclusions of Interviews

The respondents (the teachers) described the design steps they made when performing a design job. The design jobs could be an exhibition, a layout in a magazine and online education etc. The interviews reveal that the respondents have a strategy and they do not work hapharzad. Their processes resemble in such way that the respondents desire to understand for instance the users and the customer’ requirements. They deepen their knowledge about the problem and they start to generate ideas and solutions. The solutions are re-considered and the mission becomes more concrete. The respondents tend to be most creative when they have deepen their understanding and gained knowledge about the case. There has to be a thought behind an activity if anyone or anything should be considered creative (Raudsepp, E. (1993).

A behaviour scientist and a graphical designer can co-operate but conflicts are raised because of professional differences in how to deal with time and money (Nijhuis and Boersema. 1999). It is obvious that the respondents consider time as an important factor. Was it a lack of time? Did the respondent do as much as he could do with the time he had? Did he manage the time limit (deadline)? The time decides the design of the information for instance the style of an illustration. When answering the questions when the respondents are satisfied with their result and what they think is good quality, the respondents’ connect time with the quality of their work and if they are satisfied with the result.

The respondents’ satisfaction is also influenced by interplay with opinions from the customer, the user and the respondent himself. Did the customer receive what he/she wanted? Did the user understand the information? Did the respondent have enough time to perform what he wanted? There are occasions when a respondent is satisfied when the customer is satisfied. But Respondent 1 directly connects his own judgement to the customer opinion about the result. He says that when he is satisfied the customers seem to be.

The respondents have a mental image of the result. Because of compromises caused by for instance a lack of time, they know that the result could have been more successful. As respondent 3 remarks: "…often, as a designer, you know the answers to how it could have been. You knew it could have developed much more." The respondents can also be satisfied when trying something new or succeeding a project. There is an intrinsic value in performing a good job.

When answering the question if there is anything unique with an information designer the respondents says that an information designer has a consciousness way of designing, and an information designer has a holistic view and understand the requirements of the target group. An information designer can analyse a product. She/He knows why the finale product became in a specific way and why a designer works in a certain way. An information designer knows which information should be brought forward in an information set and he/she has an informative way of thinking.

The topic Information Design is unique but not its occurrence. There are professions, for instance graphical designers, working with design of information, without calling themselves information designers. Respondent 4 argues that information design is about common sense. Here is a hint of a conflict between Information Design as an academic topic (theory) and craftmanlike skills (practically). Some respondents tend to think that information design is a practice, a convention, which influences their work, without they being really consciousness about it.

8. Discussion

Research Methods in Information Design?

The study of the scientific magazines and among employees in the research team demonstrates that a multitude of research methods can be found in Information Design. Researchers from different scientific traditions are a necessity if we should be able to understand a complexity in a situation or in a problem. Since an omnipotent research method do not exist (in Information Design), a researcher should be free to use a method/methods she/he finds suitable when for instance verifying or falsifying a hypothesis. Hence it is important to establishe a terminology within Information Design, so researchers have a common language, and consequently they can interpret diverse consepts and evaluate a content/a research method.

Scientific Attitude

The study points out that the respondents (the teachers) have a structured working method when designing. An information is not a tabula rasa, an island, cut of, pure and free when designing. The respondents analyse the mission, conditions are being examined and new knowledge is established they create a framework.

According to the interviews, the framework affects the respondents. The requirements of the target group and the desires of the customers also affect the respondents. Time is a decisive important factor when seeking for new knowledge and when shaping the information set. The respondents are affected by time in such way that times can decides the artistic style and time influences the respondents' evaluation of the result.

The topic Information Design is founded on theories (based on scientific knowledge) and practical craftmenlike experience. When working scientifically with Information Design the scientific knowledge building and the handicraft should be examined. What are the similarities and what are the differences? The respondents do not use the same method to gain knowledge. Some carry out observations, some conducts interviews and some perform literature studies. The use of diverse methods is probably a result of dissimilar information sets, different target groups and different purposes. The respondents have a similar attitude – a hermeneutic attitude. They work with the whole and the parts, they review, interpret, alter and change, they review, interpret… As Respondent 3 describes that he contacts expertise, ask questions, writes new text, and alter old ones. A hermeneutic spiral.

The scientific methods used by the research team and in the scientific magazines are similar to the methods used by the respondents. The methods concerns applied research, with a majority of qualitative studies; a hermeneutic attitude is often described, and a common purpose is to understand humans' apprehensions and behaviours. As Snodgrass and Coyne (1997) we conclude that information designers, who operate both practically and with a scientific approach, require a hermeneutic attitude.

Creativity and the Design Process

The respondents tend to work in a "fluffy cloud to become more square", as respondent 5 describes it. The thought and the performance are connected to each other. The discussion group confirms and tend to believe that the thought is primary in Information Design.

The creative process and the design process are not the very same process. Creativity is a part of the design process and crops up repeatedly during the process. The respondents, as information designers, are not totally consumed by the process; instead they have a constant correlation with the context. The respondents seem highly consciousness of their relation to the framework and the working activity. Obstructer lines of action do not exclude an analytic thinking. If anyone or anything should be considered creative the activity has to be based on a thought.

Creativity and science are not opposites. Curiosity and creativity is two of the most important personal qualities for a PhD-student according to Professor Rune Pettersson[5]. The way solutions are invented is your own way, even in science. The next step is to verify the solutions as a part of scientific method.

Attitude towards the result andtowards quality

The respondents and the research team have different attitudes toward the finale result. In the research team it is common that they become satisfied if the results have good quality, in other words, if they contribute to new knowledge, verify or falsify a hypothesis. To be accepted in a scientific magazine and reviewed by experts is a stating of quality. The attitude towards the result and good quality is united.

The respondents make a difference between their satisfaction of the result and good quality. They seem to have a mental picture of the result[6] and because of compromises, for instance lack of time, the respondents know that the result could have become better. The respondents tend to find an intrinsic value in designing information, since they can be satisfied when they succeeded to perform an information set that has a "wow-effect", or if they just manage to fulfil a project.

It is probably not possible to compare the research team with the respondents, because they have different conditions. The research team do not have a customer, and they have the possibility to be reviewed by expertise within the same field of research. The respondents have a customer and there are no distinct experts reviewing the result. It is of relevance to discuss the relation between the assignor and the information designer.

Time and Money

The respondents tend to relate their result to time. They link the concepts time and quality. As an information designer you can become challengeable (time-quality relation becomes a bias).

Time and money can always affect results. Scientifically results are influenced by external factors, since there are other interests than only the scientifically ones. A research result can be affected by the status of the scientist, political context etc. (Pinch. T. 1998). Within Information Design applied research is the most common. When working scientifically we need to consider how time and money affect a certain result. It is a question of consciousness and research ethics. How applied can research in Information Design become?

Co-operation in a Design Process and a Scientific Research Process

Some scientists within design science argue that a design process is a scientific activity in itself. In this article we enlighten that a design process do not have to be a scientific activity. The respondent do not genuine work to discover new knowledge and establish new knowledge, instead the respondents found an intrinsic value in designing. Hubka and Eder (2001) discuss that if the design process should function a scientific activity within engineer science, there is a need of: "…self-motivation, openness to newer outlooks and insights, and sufficient and suitable prior knowledge." (Ibid, 4.5). If a information design process and a scientific research process should be able to co-operate, we have to conscious-raising a scientific attitude during the design activity.

Some scientist claim that design science is based up on a positivistic attitude. It is about forming a universal design philosophy, and to establish a common base in a logical positivistic tradition. With a positivistic attitude the quesions will be focus at if designers use the same methods, since the processes does not reveal the very same results (Cross. 1999). If knowledge is as Platon defined, *justified true believe*, the knowledge that there is not one result, is also a knowledge.

Let us instead look into the concept scientific design, which includes methods, in modern design practice. Scientific design is common in industrial design while for instance measuring factors as strength/firmness. In industrial design you can also

[5] Rune Pettersson, professor in Information Design at the Department of Innovation, Design and Product Development at Mälardalen University College in Eskilstuna, Sweden

[6] Cross (1996) presents a study where product designers tend to influenced the product by the first mental picture they recived, before analysing the problem or seeking for new knowledge.

study how product can communicate with a user (www.chalmers.se 040421). It is unrealisable to design decisive methods of how humans comprehend and understand information. Let us therefore look into the concept science of design, the scientific analysis of the design activities performed via scientific methods. Science of design can be linked to the pragmatically analytical hermeneutic attitude, presented in the intention model of Von Wright. He demonstration the relation between an intention, action, a situation and a goal (Von Wright. G. H. 1971). What is our intention? What situation is present? And how will a consciousness design activity effect the goal? An information designer has to be conscious about the knowledge brought forth in a specific action if she/he want to work scientifically.

A design process can be a scientific activity if a designer has a conscious scientific approach. To what extend Information Design can be science depends on our definition of science, the attitude we have, on what theories we base our knowledge, how we evaluate the results and our opinion of what knowledge is. The craftship and the theories ought to enrich each other. While knowledge theory is a theory of how reality should be apprehended and how our ideas agree with the outer reality (environment); design theory is a theory of how reality can be created and how ideas and experiences can shape an outer reality (environment).

9. References

Andersson & Elfving. (2003). *Shifting Communication Paradigm Changes Course of Product Development in Clusters*. In Griffin, B. (Eds). (2003) Selected Readings: IVLA-conferece, 2003, Changing Tides, Newport, USA.

Araï. D. (2001). *Introduktion Till Kognitiv Psykologi*. Sverige: Studentlitteratur.

Buckland. M. K. (1991). *Information as a Thing*. University of Calefornia. Berkeley. CA 94720. School of Library and Information Studies.

Chalmers University of Technology: http://nordcode.hut.fi/index.html

Cross. N (Eds). Christiaans. N (Eds). Dorst. K (Eds). (1996). *Analysing Design Activity*

Dasgupta. S. (2003*). Multidisciplinary creativity: the case of Herbert A. Simon*. Institute of Cognitive Science. University of Louisiana at Lafayette, Lafayette, USA.

Dodig-Crnkovic. G. (2003). *Theory of Science*, Course material for Philosophy of Computing and Information (VT 2004), Mälardalen University http://www.idt.mdh.se/~gdc/pi-network.htm

Goldstein. E. B. (1998) *Sensation and Perception*, New York. USA: Wadsworth Publishing Company

Floridi. L (2003). *Part I: Four Concepts*. I Blackwell Guide To The Philosophy Of Computing And Information. Oxford. UK: Blackwell Publishing

Hagoort, G. (2000). *Art Management. Entrepreneurial Style*. Delft: Eburon Publishers.

Horn, R. E. (2000). *Information Design: Emergence of a New Profession*. In Jacobson, R. (Eds). (2000). *Information Design*. Cambridge. Massachusetts: MIT

Hubka, V. & Eder, E. (1997). *Design Science*: *Introduction t the Needs, Scope and Organization of Engineering Design Knowledge*. New York: Springer-Verlag Berlin Heidelberg.

Högskoleverket. (2000). *Designutbildningar i Sverige. En utredning och utvärdering*. Stockholm: Högskoleverkets rapportserie 2000:11R.

Högskoleverkets webbplats för information om högskolestudier, *studera.nu*, http://www.studera.nu/index.shtml

International Institute for Information Design. (IIID): www.iiid.net.

Jacobson, R. (2000). *Information Design*. Cambridge. Massachusetts: MIT P

Johannessen, K. (1999). *Praxis och tyst kunnande*. Stockholm; Dialoger

Lantz. A. (1993). *Intervjumetodik : Den Professionellt Genomförda Intervjun*. Sverige: Studentlitteratur AB

Lundequist. J. (1995). *Design Och Produktutveckling : Metoder Och Begrepp*. Sverige: Studentlitteratur AB.

Nigel, C. (2002*). Designerly Ways of Knowing: Design Discipline Versus Design Science*. Data base: Direct Science. Massachusetts Institute of Technology. *Design Issues*: Volume 17, Number 3 Summer 2001. Revised version of a paper prepared for the Design + Research Symposium held at the Politecnico di Milano, Italy, May 2000.

Margolin. V. (2002): *The politics of the artificial – Essays on Design and Design Studies*. Chicago: The University of Chicago Press.

MCPee. K. (1996). *Design Theory and Software Design*. Technical Report TR 96-26. Department of Computer Science. The University of Alberta Edmonton, Alberta, Canada.

Nijhuis, W. & Boersema, T. (1999). *Cooperation between graphic designers and applied behavioural researchers*. In H. J. G. Zwaga, T. Boersema, & H. C. M. Hoonhout. (Eds.). (1999). *Visual information for everyday use. Design and research perspectives*. London: Taylor & Francis Ltd. Philadelphia: Taylor & Francis Inc.

Nylander. O. (1999). *Bostaden Som Arkitektur.* Sverige: Svensk Byggtjänst.

Pettersson, R. (2002). *Information Design an Introduction*. Amsterdam, Philadelphia: John Benjamins Publishing Company.

Pettersson, R. (2003). *Ord, bild & form – termer och begrepp inom informationsdesign.* Lund: Studentlitteratur.

Pinch. T. (1998). *Golem, What You Should Know About Science*. Cornell University. New York: Cambridge University Press.

Ramirez, J.L. (1995). *Designteori och teoridesign*. Stockholm: Nordplans Medd.

Raudsepp, E. (1993). *Kreativitetsboken.* Stockholm: Pocket express. 1993

Snoddgras & Coone. Hermeneutical? Architectural Theory Review, Journal of the Department of Architecture. The University of Sydney, Vol. 1, No. 1, pp 65-97.

von Wright, G. H. (1971). *Explanation and understanding*. Citation in K. S. Johannessen. (1997). *Tradisjoner og skoler i moderne vitenskapsfilosofi*. Bergen: Døves Trykkeri AS

Correctness Criteria For Models' Validation – The Philosophical Perspective

Irobi Ijeoma Sandra
Mälardalen University, Sweden.
ijeomasandrai@yahoo.com

ABSTRACT

Valid models are central to the existence of Computer science as in most other disciplines, but at what point can one say that a model is valid and hence correct?

A model is often taken to be an abstraction and simplification of reality (of the system being modelled) but reality (the nature of measured data, environmental and human factors) in itself, has a nature of abstract complexity, hence a 'correct' model could at best be judged as one which is *'closest'* in representation to the real system, but the question is *just exactly how close should 'closest' be to be correct?*

In this paper, we shall examine some common and general correctness criteria for models validation and seek to relate them to various philosophical perspectives to see how much information the basis of acceptance of such *valid models* could give (content and truth).

We shall also strongly explore and consider the salient philosophical angle, which presents validation *only* as a method to improve the level of confidence in a model and not a demonstration of its 'truth' content. Models should not be used as a substitute or sole basis for critical thoughts, considerations or major decisions but should be viewed just as a tool for improving judgement and intuition.

1 INTRODUCTION

Over time, simulation models have gained grounds increasingly in being used in solving problems and aiding decision making in several disciplines.

The use of a simulation model can be viewed as a surrogate for experimenting with an actual system whether it exists or is a mere proposal, which could be disruptive, not cost-effective or just impossible. Developers, users of these models and decision makers who make use of information obtained from results of these models are all concerned with whether a model and its results are correct.

However, the simulation model of any system could only be an approximation of the actual system no matter how much time or money is spent on the model building. Hence if the model produced is not a 'close' enough approximation to this actual system, conclusions derived from such model are likely to be divergent and erroneous, leading to possible costly decision mistakes been made.

More so, it is important to note that there is no such thing as a absolute model validity since a model is supposed to be a mere abstraction and simplification of reality. Therefore the definition that would be accorded model validation in the context of this paper would be that of:

"*substantiation that a computerized model within its domain of applicability possesses a satisfactory range of accuracy consistent with the intended application of the model*" [1].

From the above definition, it follows that a simulation model should always be developed for a particular set of objectives. In fact, a model which is valid for one objective may not be valid for another.

There are several 'scientific' grounds on which models are built. There are even much more techniques on which their validation rest. Philosophically, the premises on which these validations are based tend to raise more questions than answers in their forms of correctness. It is this philosophical angle that would constitute the nerve of the discussion of this paper.

The section 2 of our paper deals with validation definitions in related work and section 3, with forms of validation while section 4 handles some techniques for developing and validating models. In section 5 we shall discuss some sources of errors in models, while in section 6 we shall raise and analyse the philosophical aspects of the validation criteria in the light of information provided by the models (content and truth). In section 7 we shall point out areas of possible future work and then conclude in section 8.

2 DEFINING VALIDATION

Much research work has been done with respect to simulation models.

In their paper, Robert G. Sargent et al states three basic approaches used in deciding whether a simulation model is valid or invalid. These approaches are:

- The development team takes the decision as to whether the model in question is valid. This is a subjective decision based on the outcome of different tests and evaluations done as part of the model development process.

- The use of a third party to decide if the model is valid, independent of the model developers and users. This method is often used when cost associated with the problem the simulation model is needed to address is high and also in terms of certification of credibility.
- The use of scoring models in which scores are determined subjectively when conducting various aspects of the validation process and then added together to determine the category/overall scores for the simulation model. In this case, a simulation model would be considered valid if its total scores and category scores are higher than the passing scores.

Averill M. Law et al [2] reinstates that validation can be done for all models regardless of whether their corresponding systems exists presently or would be built in future.

Also in their paper [3], Jack P.C. Kleijnen et al give insight on validation of models using statistical techniques and reasoned that the technique that should be applied would depend on the availability of data in the real system. Regarding this data availability, they distinguish three scenarios namely:

- No real-life data available
- There is only data on the real output (not the corresponding input or scenario)
- Besides the output data, the corresponding input is also know

They agreed that in the event that no real-life data is available, strong validation claims remain impossible! In this case then, *sensitivity analysis could be used to support validation, which can be defined as a systematic investigation of the reaction of the simulation responses to extreme values of models' input or to drastic changes in the models' structure.*

However, these kinds of analyses show if *factors* have effects that agree with experts' prior substantial knowledge. Unfortunately, in actual practice, it is not all simulation models that have effects with known signs; still many models do have *factors* with known factors.

Kleijnen et al defined their problem entity as the system (real or proposed) that is to be modelled. The conceptual model would be the mathematical/logical representation of this entity for a given study, while the computer model is that obtained through a computer programming and implementation phase. Inferences and conclusions are therefore drawn by conducting experiments on the computerized model.

3 FORMS OF VALIDATION

There are many forms of validation. It could be seen that validation of *conceptual models* is determining that the theories and assumptions underlying the conceptual models are correct and its representation of the problem entity reasonable for a given purpose. The big question here is whether the conceptual model contains all the necessary details to meet the given objectives.

The *operational validity* is referred to as ensuring that the model's output behaviour has enough accuracy for its intended purpose on its domain of applicability whereas data validity is defined as determining that the necessary data for model construction, evaluation and testing are adequate and correct.

Data validation deals with determining that the data needed for building the model, experimentation and validation are adequately sufficient.

White-box validation is the process of determining that constituent parts of the model represent the corresponding real-world elements with adequate accuracy. The big question here is whether each part of the model represents the real world with enough accuracy.

The *Black-box* validation is concerned with determining that the total (entire) model is an adequately accurate representation of the real world.

Several validation techniques are used. According to [4] there is *no algorithm* or particular pattern to select a given technique to use. However, there are several factors which affect the choice of the techniques that one would use.

4 VALIDATION TECHNIQUES

In this section we shall look at several validation techniques often used, and we shall later consider their philosophical interpretations.

<u>Comparison to other models:</u> Different outputs of the simulation model being validated are compared to those of other 'valid' models

<u>Degenerate test:</u> This has to do with appropriately selecting values of the input and internal parameters to test the degeneracy of the model's behaviour. For instance, to test to see if the average number in the queue of a single server continue to increase with respect to time when the arriving rate is larger than the service rate.

<u>Events validity:</u> The events of occurrences of the simulation model are compared to those of the real system to see if they are similar.

Face validity: This is often used to know if the logic used in the conceptual model is correct and if the input-output relationship is reasonable. This has to do with asking knowledgeable people if the system model behaviour is reasonable.

Historical Data validation: If data was collected on a system for building or testing the model, part of the data are used to build the model and the remaining data are used to test if the model behaves in the same way the system does.

Predictive validation: Here the model is used to forecast the system's behaviour and the model's forecast to determine if they are the same.

Traces: Specific entities in the model are followed through the model to know if the logic of the model is correct and if the necessary accuracy is obtained.

'Turing tests': Knowledgeable experts on the system are asked if they can differentiate between the output of the system and model.

Schellenberger's Criteria: This include *technical* validation which has to do with identifying all divergences between the model assumptions and perceived reality as well as the validity of the data used, *operational* validity which addresses the question of how important these divergences are and dynamic validation which ensures that the model will continue being valid during its lifetime.

5 SOURCES OF ERRORS IN MODELS

These may not be independent.

- *Model-structure*. In both the conceptual model and the mathematical model important physical phenomena might be omitted or overlooked, and mathematical simplifications might be inadequate for capturing complex dynamics.
- *Numerical solution*. The solution of the numerical model might differ dramatically from the (unknown) ideal solution of the mathematical model.
- *Calibration*. Residual uncertainty about values of model parameters remains after calibration.
- *Input values*. Proper numerical values of the code inputs that describe the scenario for prediction might be known only approximately.

6 PHILOSOPHICAL PERSPECTIVES ON MODELS VALIDATION

Considering the questions: '*what does it mean to validate concepts? Or what are the criteria? Both philosophers and scientists have been unable to agree about the answers to them.* [Adapted from Shannon, 1975, p. 211]'.

In this section, we shall examine crucial questions arising from the validation criteria of models that have been mentioned above.

The computer science (or information science in general) is faced with this difficulty more so than social sciences because of its diverse constituents, ever-changing contextual environment (technology), and relatively short life span.

Validation assures that a model (or each construct in a conceptual model) contains the features imputed to it in their individual definitions or description. In other words, validity implies that it is well-grounded, sound or capable of being justified.

The response of a computer science empiricist to the question "How do we validate?" could be to design *an experiment or build a prototype and test your concept or conceptual model*. But, a fundamental problem with this approach, notwithstanding the assumptions inherent in statistical experimental design, is the presupposition of the "validity" of a concept or conceptual model. That is, a belief in the notion that mere definition implies that a concept has "face validity." If simply using a "term" made it acceptable to a discipline, one would never reach an agreement on commonly held truisms or knowledge of that discipline.

Simulation models are believed across disciples to give *information* on the real system. In [5], a 21st century philosopher Luciano Floridi defines information as basically comprising 'content and truth'.

In philosophy, there is a huge difference between truth and correctness. While truth is an absolute, correctness is relative to the system. For example, if you read a book on the philosophy of mathematics, "truth" is not the issue because mathematics does not deal in truth but deals with provability. Maybe physics deals in truth, because the job of science and engineering is to understand the world as it is. Thus the issue for consideration here is *correctness*.
An important question to ask in this context would be: *can simulation models yield knowledge about the real world?*

The epistemological importance of this question is such that if the answer is *no*, then what many scientists are doing nowadays is just playing with computers, not creating new knowledge!

However, considering the practical importance of the question, if no is still the answer, it means that the several policies which are now based on simulation models would grossly be misguided. It is interesting to note however that even in the field of philosophy, varying opinions do exist about whether verification and validation are possible or not.

In [6], an interesting philosophical argument issues between Oreskes et al and Fredrik Suppe in trying to proffer solution to this seeming deadlock.

Oreskes et al strongly argues that simulation models cannot be verified and hence scientists cannot obtain knowledge from simulation modelling. On the contrary, Fredrik Suppe retorts that simulation models can be verified in some sense and hence knowledge could be obtained from them. Some important issues that readily comes to mind in this case would be a deep consideration of some epistemological questions such as

- What (and how) do we learn from experience?
- What is the correct way of learning from experience?

There are also several traditional philosophical views, which include Inductivism (enumerative induction, inference to the best explanation and Bayesianism) and Falsificationism.

However, Oreskes et al argues the above, utilizing traditional philosophical debate over inductivism. Their criticism of the traditional view in three different areas stemmed from Hume's problem of induction, which says that

- All inductive reasonings are based on the assumption of uniformity: What we have observed and what we haven't yet are basically similar. According to him, the question would be: 'why can we rely on such an assumption?' Nothing we have observed until today does not assure that the same regularity will hold tomorrow (unless we use induction --- this is a circular argument).

- **Underdetermination**

  - Given any amount of evidence, there are mutually incompatible theories which equally fit with the evidence
  - when a prediction from a theory contradicts with the observation; there are various mutually incompatible ways for making the theory compatible with the evidence.

- **Theory-ladenness of observation**

  These philosophical views presuppose that our observation is somewhat independent from our scientific theory. But what we see is strongly influenced by our background knowledge and assumptions. A common example would be asking a zoologist and a social scientist to give interpretations of a diagram of a rabbit.

Why do we care about theory-ladenness of observation? This is because a conflict between two incompatible theories is supposed to be settled by doing some experiment or observation. However, Theory-ladenness can cause a serious problem with such a procedure.

Considering the Underdetermination vs. Theory-ladenness, the difference between the underdetermination thesis and theory-ladenness can be summarized as follows:

Underdetermination
Same evidence -> Incompatible theories

Theory-ladenness
Incompatible theories -> Different evidence.

In the actual sense, arguments by Oreskes et al. are an application of these traditional criticisms of induction to simulation models.

6.1 Degrees of certainty

However, an interesting categorization was projected by Oreskes et al in which they made the following distinctions: They inferred that there were various degrees of certainty:

- Absolutely true (logical truth) ie verification
- Plausible, probable (in terms of evidence) > confirmation
- Consistent (not contradictory) > validation

Therefore from the philosophical analogies given above it can be deduced that:

(a) Models *cannot* be verified in that there is no logical proof that a model is true.

(b) Models *can* be validated, this means that we can prove that a model does not contain a detectable flaw and thus internally consistent.

This can be evident in:

- Comparisons:

  If two totally different ways of solving the same problem give the same answer, these ways of solution may be reliable.

- Calibration:

  Adjust initial values so that the model can accommodate known data. These procedures are far from *verifying* the model.

- Confirmation:
  Models may yield predictions that match with observation, but this means only that the model is probable, not that the model is true.

Therefore from the above analysis, Oreskes et al concludes that:

- The primary value of a simulation model is heuristic, that is, to give evidence to strengthen what may already have been partially established through other means, for instance, sensitivity analysis, or even challenging existing formulations.
- A simulation model is a `fiction'. It is never a `real thing'. (Cartwright).

In contrast to the above views, Suppe assumes a less strict philosophical stance as follows:

(1) It is true that we cannot logically prove that a model is true. But maybe their way of defining 'verify' is too strict. Do we really want that absolute certainty? That makes all empirical knowledge impossible.

(2) Extra factors can affect the result. But still a simulation model is creating knowledge about the real world when the system is isolated or other factors are negligible.
(3) Don't take underdetermination too seriously. Often it is hard to find even one reasonable solution.
(4) Don't take assumption-ladenness of simulation models too seriously, either.
(5) An important aspect of modelling is the mapping relationship between three systems. As far as this mapping relation holds, a simulation model is a representation of that aspect of the real world, not just a heuristic tool.

With view to the above two major open and highly contestable areas, one could strike some good balance by answering the following questions:

- What level of certainty do we want for scientific knowledge?
- Can simulation models provide that level of certainty?

6.2 Possible integrations?

In [7] Khazanchi attempts to integrate notions from the philosophy of social sciences, the information systems (IS) field and its referent disciplines and sets forth a framework for the validation of IS concepts. The proposed philosophical framework for validation of concepts and conceptual models consists of a set of "criteria for validation" of concepts.

He asserts that as a concept fulfils each succeeding criteria its potential ability to have inherent "truth content" with regard to its general acceptance in the field strengthens. After all, "... concept formation and theory formation in science go hand in hand.... The better our concepts, the better the theory we can formulate with them, and in turn, the better the concepts available for the next improved theory." [Paraphrased from Kaplan, 1964, p. 52-54].

The following are his suggested criteria for such validation:

1. Is it plausible? A concept or conceptual model is plausible if it has face validity. Plausibility establishes that this model is more than just a belief. This criterion is useful to assess the apparent reasonableness of an idea and could be demonstrated by deduction from past research or theories, or, it could be developed on the basis of observation or induction.

2. Is it feasible? This criterion dictates that a concept or conceptual model, at the least, has the quality of being workable. Added to plausibility, a feasible concept or conceptual model would be operational in that it would be amenable to verbal, graphical, mathematical, illustrative, prototypical characterization.

3. Is it effective? This criterion deals with the question: How effectively does the model describe the phenomena under study? Also an effective concept or conceptual model has the potential of serving our scientific purposes [Kaplan, 1964]. It also guides and stimulates other scientific inquiries.

4. Is it pragmatic? The pragmatism criterion dictates that a concept or conceptual model should not be restrictive to the extent of logically excluding previously valid models. Thus, this criterion provides that concepts or conceptual models should subsume, for obviously practical reasons, any conceptual structures that previously explained related phenomenon. Hunt [1990] demonstrates this criterion with the example of Newton's law. He argues that simple pragmatism would require that any new conceptual development could not preclude Newton's laws (as in the case of Relativity, where these laws are a special case subsumed within relativity). In effect this criterion emphasizes that concepts and conceptual models should have some degree of abstract, logical self-consistency or coherence with other concepts and conceptual models in the discipline.

5. Is it empirical? (Does it have empirical content?) Empirical content implies that a concept or conceptual model must have "empirical testability" [Hunt, 1990]. In this vein, Dewey also affirms that although concepts can be developed without reference to direct observation, and although this logical conceptual development is indispensable to the growth of science, the ultimate test of a concept or conceptual model lies in having the ability to

empirically collect data to "corroborate" it. According to Dewey [1933, p. 183], "Elaboration by reasoning may make a suggested idea very rich and very plausible, but it will not settle the validity of that idea.

6. Is it predictive? (Does it explain a phenomenon that is expected to occur?) We can better understand the meaning of this criterion through words of Rashevsky (1954, p. 152-3): "A theory or theoretical concept is considered the more convenient or useful, the better it enables us to predict facts that hitherto have not been observed... The scientist constructs theories, theoretical concepts or theoretical frames of reference that are isomorphic with the world of observable phenomena. This isomorphism is never complete, never covers the whole range of observable phenomena... wider the range of isomorphism, the greater predictive value of the theory." Thus, a concept or conceptual model that is predictive would, at the least, demonstrate that given certain antecedent conditions, the corresponding phenomenon was somehow expected to occur [Hunt, 1990].

7. Is it intersubjectively certifiable? Hunt [1990], Nagel [1979], and several others are of the opinion that all scientific knowledge, and in consequence, concepts or conceptual models "must be objective in the sense of being *intersubjectively certifiable*." This criterion provides that concepts or conceptual models must be "testable by different investigators (thus inter-subject)." Investigators with differing philosophical stance must be able to verify the imputed truth content of these concepts or conceptual structures through observation, logical evaluation, or experimentation.

8. Is it intermethodologically certifiable? In addition to being intersubjectively certifiable, this related criterion provides that investigators using different research methodologies must be able to test the veracity of the concept or conceptual model and predict the occurrence of the same phenomenon.

7 FUTURE WORK

The questions that have been explored in this paper have by no means conclusive answers.

A good future work area would be to view a models' validity philosophically as a measure of the model's *absolute* truth content, not just theoretically but experimentally as well.

8 CONCLUSIONS

We have explored validation of models and the general criteria on which it is based. We have also considered common techniques available for these validation and common error prone areas.

Most importantly, we were able to look into several burning philosophical issues, views and opinions held in this area and have come to the vital conclusion that even though models' verification is still highly contestable, model validation is seen even philosophically as a 'can-do'. The eight Khazanchi's postulated criteria further gave insight as to how to 'test' a models' inherent 'truth content'.

REFERENCES

[1] Robert G. Sargent: Verification and Validation of Simulation models

[2] Averill M. Law, Michael G. McComas: How to Build Valid and Credible Models – *proceedings of the 2001 Winter Simulation Conference*

[3] Jack P. C Kleijnen: Validation of Models: Statistical Techniques and Data Availability – *Proceedings of the 1999 Winter Simulation Conference. P. A. Farrington, H. B. Nembhard, D. T. Sturrock and G. W Evans, eds.*

[4] Stewart Robinson, Simulation Model Verification and Validation: Increasing the User's Confidence – *Proceedings of the 1997 Winter Simulation Conference, ed. S. Andradottir, K. J. Healy, D. H. Withers, and B. L. Nelson.*

[5] Luciano Floridi: Philosophy of Information as a new field of Research.

[6] Suppe vs. Oreskes et al. Verification, Validation, and Confirmation of Simulation Models: May 7, 1998 PHIL 250

[7] Deepak Khazanki, A Philosophical Framework for the Validation of Information Systems Concepts.

Other citations included directly are taken from:

Dewey, J. How We Think, Boston: D. C. Heath & Co., 1910 (1933 reprint).

Hunt, S., Marketing theory: The philosophy of marketing science, Homewood, IL: Richard D. Irwin, Inc, 1990.

Kaplan, A. The conduct of inquiry, Scranton, PA: Chandler Publ. Co., 1964.

Philosophical aspects of formal verification

Oskar Wibling

***Abstract*— Formal verification of computer systems is the art of mathematically proving that a software system or model of a system conforms to its specification [1].**

In this paper we illuminate a number of philosophical issues that arise in connection to this practice. These issues regard expectations, the amount of abstraction in system models, reasons for differences between a requirement specification and the actual system implementation, as well as the role of logics. We then identify a number of formal verification catalysts and impediments. Our conclusion is that regardless of whether expectations on formal methods are too high or too low, their most important quality is that they can improve the reliability of software. If they will, in the end, succeed in providing the help that is required is still an open question. For future systems to provide the level of reliability that is expected and needed, however, the success of these methods will be a necessity.

***Index Terms*— Formal verification, logics, modeling, philosophical, ethical issues**

I. Introduction

As computer systems get larger and more complex it becomes increasingly difficult to ensure that they work as intended. Furthermore, if the system under consideration is a safety critical one its correct operation is of great importance to many people. Examples of such systems can be life support systems, transportation systems as well as other technologies that in case of malfunction can endanger people's lives.

In the computer program development industry of today, the most common ways of trying to certify a system's correctness are to either use simulations, employ other types of testing, or do both of the above. These approaches fail to provide conclusive guarantees, however, since they only try out a limited number of scenarios as opposed to all of them. There will thus always be cases left that have not been tried or verified.

To remedy this problem, there has been extensive research over the last decades [8], [26], [28] aiming at developing methods by which one can mathematically prove that a system behaves correctly. If the system, or rather the system model, passes the test then it is expected to work in all situations and not just for a limited few.

In this paper we survey philosophical issues that are important to consider in connection to the field of formal methods. The remainder of the paper is organized as follows. Section II discusses current expectations on formal methods. Section III briefly covers the level of abstraction in models. Sections IV and V describe the role of requirement specifications and logics respectively. Sections VI and VII respectively list and comment on potential formal verification catalysts and impediments. Finally, Section VIII provides our conclusions.

II. Expectations on formal methods

As one can imagine, the mathematical proofs of correctness rapidly tend to become complex for larger programs. A manual proof quickly becomes a very tedious job to construct. Even though there are automatic tools available it is at the moment not feasible to verify all aspects of large "real life" systems.

There is currently a lot of research being conducted with the goal to make system verification more feasible, whereby there is also hope for greater acceptance and utilization in the software development community. To this extent, several alternatives have been proposed, ranging from formal construction (code refinement) methods to modularization approaches where each part of a system is verified in isolation after which the complete system can be checked under the assumption that the individual components work as intended.

Even so, the majority of system developers still resort to common inductive testing. The main reason for this is presumably ignorance; in general there is simply no knowledge that there exists an alternative to the experimental approach. The solution to that problem could be better education, something which has already started to happen.

However, even among the enlightened few, there is a large group of opponents to formal methods. Expectations are often too low; it is generally believed that these methods are difficult to apply and that the return will be of little value. The opposite, too high expectations, is perhaps not as common but something which could be just as dangerous.

Since a computer system in the end carries out its work by means of physical processes (the electronics that

Oskar Wibling is a member of the Mobility and Communications Research Groups of Uppsala University, Sweden. E-mail: oskarw@it.uu.se .

[1]Or, alternatively, the validation of a specification against a list of (formalized) requirements on that system.

constitute a computer) it is prone to errors not possible to foresee in the abstract system model or even in the source code itself [13]. Wherever one draws the line for verification, be it at the level of the algorithms used, the program, the assembly code into which the program is compiled or even at the physical level one can never be completely sure that the system will work as intended. There is always the potential of a bit flip triggered by a random error in one of the physical components used [15]. This can for example occur as the result of the influence of terrestrial cosmic rays. Even when using hardware error correction, a 1 Gb memory system based on 64 Mbit DRAM:s has a potential for about 900 errors in 10000 machines in 3 years [20]. Hence the notions of relative versus absolute correctness [9].

Furthermore, even if a system has been verified, how can one be sure that the verification tool, a piece of software itself, does not contain errors which prevent correct verification? Fetzer has said that [13]:

> "...the theoretical possibility of subjecting [theorems and programs] to rigorous appraisal ought to be regarded as more important than its actual exercise."

Even if one agrees on this ranking of importance, it can be regarded the duty of every system developer to at least strive towards perfection. Methodologies should be improved as much as possible. As computer systems become more and more widely used, already influencing every aspect of our everyday lives, it can certainly be regarded an ethical issue that we should do the utmost to make them as reliable as possible. Safety critical systems such as nuclear power plants or aeroplanes are of course the most prominent and evident examples.

Fetzer, in the discussion [12] following his original article [13] reformulates his point as follows:

> "[...] since program verification cannot guarantee the performance of any program, it should not be pursued in the false belief that it can - which, indeed, might be entertained in turn as the 'ill-informed, irresponsible, and dangerous' dogma that my paper was intended to expose."

Charlie Martin in the same discussion [12] responds to this by saying that:

> "However, MY point, and the point made by many others who have responded to Dr Fetzer, is that program verification can provide a sort of *relative* guarantee: the likelihood of defects in programs constructed with proofs is very much smaller."

He continues:

> "Furthermore, this relative guarantee is the ONLY kind of guarantee that ANY engineering technique [...] can provide. All physics and all engineering is based on the implicit or explicit assumption that the mathematical model, the abstraction, that is manipulated is sufficiently close to the behavior of the real world."

Will it be possible in the future to build a perfect machine in which hardware errors cannot occur? It does not seem a realistic thought in a chaotic universe, yet one cannot be completely sure of this. As long as this theoretical possibility exists we cannot either rule out the utopia of a completely verified computer system.

However, until then, our expectations on formal methods should not be higher than those on any other engineering methodology. These methods (usually) improve the reliability of a system and this is a very attractive quality indeed.

III. The distinction between model and reality

Charlie Martin says in the discussion [12] following Fetzer's article [13] that

> "...one can argue on the same basis that euclidian geometry has no use and gives no insight, because the same distinction between mathematical model and physical world exists; worse, Euclid's axioms for plane geometry can be demonstrated false in the real world. Thus it is useless and morally wrong for surveyors to learn euclidian plane geometry, because the real world doesn't fit the model."

This is an example which illuminates the very core of the problem.

There are approaches in which a model of a system is not used for the verification but rather the implementation, i.e. the source code itself [21]. In those cases, however, other limitations regarding what to focus on in the verification have to be imposed in order not to run into overwhelming complexity.

In any case, the result is almost always at this point that the model has to be an abstraction of the real system. The important part is that the model reflects all the relevant properties in the sense that both the system model and the system itself conform to the requirement specification.

IV. The importance of the requirement specification

A. *What's in a specification?*

According to Steve Savitzky in the discussion [12] following Fetzer's original article [13], the real problem with verification is that it gives a false sense of security because:

1) "You don't know whether what you initially specified is what you really wanted."
2) "You don't know what real requirements that were left out. These can be reasonable bounds on time, memory, accuracy, etc."

Furthermore, a large portion of the errors in computer systems are bound to be caused in the glitch between the requirement specification and the system implementation, i.e. in the interpretation of the specification. If this is done by a human being, which is commonly the case, then the individual knowledge and experiences come into play.

Is the future of formal methods then rather in the pursuit of applicable formal construction methods? This would at least circumvent the error prone step between the specification and the implementation.

Automatic program construction methods, aimed at for example the making of components that parse XML [27] files for insertion into a database, are already employed in various software projects [16]. UML [25] is also used in several design tools to formally model systems using the specification as a basis. Then code skeletons such as Java classes can be automatically generated from the model. This also reduces the possibility of errors in the modeling phase since the UML model can be validated towards the specification prior to generating the code.

What then seems to be most important in the end is that the specification really portrays the requirements that the person (or group of persons) ordering the system had intended. Furthermore, even if this is the case, it is very likely that the person ordering the system did not have a clear view of what they wanted and therefore added some fuzzy and imprecise requirements. Often, the initial view of the requested functionality is vague and the system (in the best case) is a result of an interaction between the one who is ordering and the one who is designing/constructing it. This is a dynamic process in which premises and conclusions are changed many times (looping/iterative).

The problem of imprecise requirements can be reduced if some kind of (monotonic) logic is used for the specification. Then, contradictions and vague statements will be impossible. Even then, though, some requirement can accidently have been left out.

NASA's Langley formal methods team [22] mention three "basic strategies that are advocated for dealing with the design error":

1) Testing ("Lots of it")
2) Design diversity, i.e. software fault-tolerance: N-version programming, recovery blocks, etc.
3) Fault avoidance, i.e. formal specification/verification, automatic program synthesis, reusable modules.

In conclusion, the correctness of the requirement specification is vital for the correctness of the finished system and all measures practically possible should be taken to ensure this.

B. *Formal construction methods vs rapid prototyping*

An alternative to first developing a system and then attempting to verify it, is to employ a formal construction method [1]. This means that the formal specification is used as a basis and through a number of refinement steps a complete system model is generated. As long as the steps of the construction process have themselves been verified, the resulting system model does not have to be further checked. It is guaranteed to satisfy the requirements directly. From this refined model, code can then be automatically generated whereby the system model actually becomes the executable system. Naturally, this translation (or compilation) has to be verified as correct, but this presumably only has to be done once. The issue still remains, however, about how to make sure that the requirement specification truly reflects the customer's wishes. One has to assume correctness at one stage and at the point of the requirement specification this is commonly done by manual inspection and interaction with the person ordering the system. An alternative to presenting the customer with a list of formalized requirements, although not as formal, is to use rapid prototyping. One can also imagine a combination in which one or more prototypes are first used when discussing requirements with the client. These are then used as a basis for formulating a series of more formal requirements.

Fred Brooks [5] has written about automatic programming, requirements refinement and rapid prototyping. What he refers to as "automatic programming" is "[...] the generation of a program for solving a problem from a statement of the problem specifications". Regarding rapid prototyping he writes:

> "The hardest single part of building a software system is deciding precisely what to build. No other part of the conceptual work is as difficult as establishing the detailed technical requirements, including all the interfaces to people, to machines, and to other software systems. No other part of the work so cripples the resulting system if done wrong. No other part is more difficult to rectify later. [...] Therefore, one of the most promising of the current technological efforts, and one that attacks the essence, not the accidents, of the software problem, is the

development of approaches and tools for rapid prototyping of systems as prototyping is part of the iterative specification of requirements."

V. The role of logics

A. What do logical statements tell us of the system?

Most available formal methods are based on the use of some logic in which system requirements can be specified. Examples of such logics common in formal verification are classical first order logic, modal logics and flavors of the latter, such as linear temporal logic (LTL) and computation tree logic (CTL).

Those mentioned above are all monotonic logics, meaning that an addition of premises does not reduce the acceptability of the conclusion [17]. In contrast, in nonmonotonic logic [17] a conclusion that is warranted at one time may come to be rejected at a later time, in the light of new and better evidence. Furthermore, in the light of even more evidence, the original conclusion may once again come to hold, which is a consequence of the fact that it in general can be changed.

Examples of things that can be said (true propositions) about a system using a classic monotonic logic are the following (expressed informally):

- There is no deadlock in the system.
- If an output to Printer 1 has been done, then an output to Printer 2 will eventually also be done.

These two demands do not contradict each other in isolation from other propositions. There can exist a system in which there are no deadlocks and the output to one printer means that there will also be an output on a second printer. If our requirement specification would consist of these two logical statements, then if the system model could either deadlock or output to Printer 1 without at some later stage outputting on Printer 2, we would like our formal verification method to report a failure in conformance.

Bremer [3] discusses the question of whether or not logical truths carry information. What he refers to by this is the question about whether new information is gained by drawing inferences or arriving at some theorems.

> "[...] the formal accounts of information and information content which are most widely known today say that logical truth carr[ies] no information at all."

We can conclude that system requirements as exemplified above do provide us with one important additional piece of information, namely whether or not they are conformant with each other. That, however, is not the same as being able to draw some additional information from the inference of the statements.

B. What cannot be expressed in monotonic logics?

In contrast, an example of a sequence of statements, symbolizing the inductive retrieval of information that one would need a nonmonotonic logic to be able to express is the following (taken from Antonelli [2], [24]):

1) (No mammals fly)
2) (A bat is a mammal $\wedge$ a bat flies) $\Rightarrow$ The first proposition is modified to (No mammals *except bats* fly)
3) (A baby bat is a bat $\wedge$ No baby bats fly) $\Rightarrow$ The first proposition is again modified, this time to (No mammals except bats *that are not babies* fly)
4) (Stellaluna is a baby bat) $\Rightarrow$ We can infer that Stellaluna does not fly

In his paper "Real Logic is Nonmonotonic" Kyburg [17] comes to the conclusion that a nonmonotonic logic is possible in the sense that a framework for nonmonotonic reasoning can be constructed with the ability to draw conclusions that go beyond the initial premises.

According to Antonelli [2], [24], there are three major issues connected with the development of logical frameworks that can adequately represent defeasible reasoning:

1) Material adequacy; how broad a range of examples is captured by the framework?
2) Formal properties; to what degree does the framework allow for a relation of logical consequence that satisfies the conditions of Supraclassicality, Reflexivity, Cut, and Cautious Monotony?
3) Complexity; what is the computational complexity of the most basic questions concerning the framework?

Antonelli further says that:

> "There is a potential tension between (1) and (2): the desire to capture a broad range of intuitions can lead to ad hoc solutions that can sometimes undermine the desirable formal properties of the framework."

Furthermore, according to Antonelli, the development of non-monotonic logics (and related formalisms) has in general been driven, since its inception, by consideration (1). It has thereby relied on a rich and well-chosen array of examples which raises the question about if it is possible to construct a single universal framework that will inhibit the necessary properties.

Antonelli continues:

> "More recently, researchers have started paying attention to consideration (2), looking at the extent to which non-monotonic logics have generated well-behaved relations of logical consequence."

Practitioners of the field have reportedly [2], [24] so far encountered mixed success in their attempts to develop such a framework.

Regarding the connection to formal methods, in what situations one would need the help of a nonmonotonic logic to be able to express the requirements of a system is difficult to say. Nonmonotonic logics have often been studied in connection with artificial intelligence [2], [24] and it may be there we can find the best suited applications. If one would like to model a brain by using e.g. some process algebra and then specify conditions on it in a logic, then it could very well be necessary to have a nonmonotonic means of doing this. In formal methods of today, however, one initially has all the premises available and they are checked "once and for all" or, alternatively, used to formally construct a system.

One could imagine the common scenario in which a requirement specification is initially given and a system is formally constructed based on it. Then at some stage during the development phase or even after the system has been constructed and deployed, a customer may come up with some additional requirements. If these conflict with any of the original requirements, then one has to make a selection as to which requirement to favor. Here we could very well find an interesting area of application.

VI. Formal verification catalysts

A. Customer fault tolerance

Today's "consumer" of a software system is relatively tolerant regarding the reliability. If a certain program crashes once in a while and has to be restarted this is usually tolerated, almost expected. It is not until there is a rather high frequency of errors that a consumer will complain. Recently, however, voices have been raised to make software producers take greater responsibility for the programs released. Brad Cox [11] has suggested an interesting payment model in which the customer pays for their utilization of a program. The program pays for the use of its licensed components and so on. He calls this free flow of programs where the user is only charged for actual usage "superdistribution".

B. Using standard components

One central idea (for example advocated by Cox [11]) that could help to improve the quality of software is to use standard components more instead of building so much from scratch. Then formal verification could be performed "once and for all" on these components and it wouldn't matter so much that the methods are often very memory intensive or require a lot of manual assistance. This is actually the advantage of formal construction applied on the macro level. The thought is not new and is central in e.g. the Component Object Model (COM) [6].

David Bridgeland [4] summarizes Cox's claim in four points:

1) "The reason that software is costly, of low quality, and difficult to construct is that we build it rather than assemble it from prebuilt components, the way that every other engineered product is constructed."
2) "The reason we build rather than assemble is that there is not a robust market for buying and selling components."
3) "The reason there is not a robust market for components is that there is no standard mechanism for pay-per-use of components."
4) "The reason there is no standard mechanism has to do with the difference between information and atoms."

Actually, these statements are not completely true. Components can still not be charged for on a pay-per-use level, but there are more or less robust markets for them such as *ComponentSource* [10]. On these markets, system builders are charged for a binary copy of the component and can later be charged again when they want to download an updated/improved version or, alternatively, after using the component for some period of time.

An especially interesting point, however, is the last one although not well-formulated as it stands. What is meant is presumably the difference between a program component, or in fact a segment of code running on a machine, and an atom. Nicholas Negroponte [23] has written about the important difference between "bits and atoms" in the economy of today and in the future. The issue of specifying a value for bits of data, be it in the form of software components, music files, or books is a task that has yet to be satisfactorily solved.

C. Computer viruses driving correctness

Computer viruses [14] containing self-modifying code are already available and it is very likely that they will evolve even more in the future. By a program modifying its own code before each new attack it becomes practically impossible to detect its signature (or pattern). Therefore it will be more and more important to make sure that operating systems and other pieces of software are virtually free of holes.

It is plausible to construct programs against viruses in analogy with the fact that we as organisms are also prone to virus attacks but helped by our immune system if properly trained and activated. There have e.g. been suggestions for a new Internet architecture [7] in which the network is cognitive and adaptive. One can imagine virtual "battles" between metamorphic, self-learning viruses or worms and the "intelligent" network or operating systems' components.

VII. Potential formal method impediments

A. Complexity of programs defeating verification?

Fowler [14] writes that:

> "TOTALLY securing an operating system - any operating system, but particularly Microsoft Windows - is incredibly challenging."

This is no underestimation, and Microsoft [19] is also an example of a company that is very active in the research and development of formal verification methods.

When verifying a system using reachability analysis, what one essentially does is to explore all the possible states that it can reach for any given set of inputs. Control structures in a program cause branching whereby the number of states rapidly increases. In order for a verification algorithm to detect loops in the state space, the searched states need to be stored in memory so that already searched paths can be avoided. Even if a compact (often compressed) format is used for this purpose, the amount of available memory imposes limitations on the programs that can be verified. When the state space increases very rapidly, this is referred to as the "state space explosion" problem.

Several methods have been proposed to minimize the impact of the state space explosion problem [18]. However, the fact remains that it imposes significant limits on the complexity of systems that are to be formally verified.

B. Difficulty of usage

The ease of use of a particular engineering method should not be underestimated in importance when it comes to the choice of method at the end of the day. Current formal verification methods often require the user to have a significant amount of knowledge of their functionality. If this threshold could be lowered by making the verification process simpler to apply it would certainly promote the use of such methods. This is a real challenge for the formal methods community and one which may very well prove to be the most important in the end.

VIII. Conclusions

In this paper we have discussed a number of philosophical issues connected to formal verification. We can conclude that expectations on these types of methods are commonly either too high or too low, in both cases to a large extent presumably because of ignorance. Formal methods can improve the reliability of software and that is a very important quality. However, the methods of today need to be further developed if their use is to be increased. This may involve the development of new logics in which system properties can be stated. Moreover, it will involve the development of new approaches which require less expertise by the intended user community. Formal specification methods have a central and important role regarding the minimization of differences between customer requirements and system specifications. Formal construction methods can further help to make sure that the finished system is not missing any of the required properties.

In addition to lowering the threshold of use for formal methods, which can be regarded as a top down approach, we have also identified driving forces that will promote the use of such methods from the bottom up. These include the ever increasing complexity of systems as well as extended threats from new forms of computer viruses. Whether or not formal methods will succeed in the end is a question that only the future can reveal the answer to. If they fail to provide the help that is required, however, the outlook for future systems' correctness is not a bright one.

References

[1] Jean-Raymond Abrial, *Event Based Sequential Program Development: Application to Constructing a Pointer Program*, Proc. International Symposium of Formal Methods Europe, Pisa, Italy, Sept. 2003.

[2] G. Aldo Antonelli, *Logic* in The Blackwell Guide to the Philosophy of Computing and Information (L. Floridi ed.), Blackwell Publishing, 2004.

[3] Manuel E. Bremer, *Do Logical Truths Carry Information?*, Minds and Machines 13(4), pp. 567-575, Nov. 2003.

[4] David Bridgeland, *Review of Brad J. Cox book "Superdistribution: Objects As Property on the Electronic Frontier"*, `http://authors.booksunderreview.com/C/Cox,_Brad_J./`.

[5] Fred Brooks, *No Silver Bullet; Essence vs Accidents of Software Engineering*, IEEE Computer Magazine, April 1987.

[6] Carnegie Mellon Software Engineering Institute, *Component Object Model (COM), DCOM, and Related Capabilities*, `http://www.sei.cmu.edu/str/descriptions/com.html`.

[7] David D. Clark, Craig Partridge, J. Christopher Ramming and John T. Wroclawski, *A Knowledge Plane for the Internet*, Proc. Applications, technologies, architectures, and protocols for computer communication, ACM SIGCOMM, Karlsruhe, 2003.

[8] Edmund M. Clarke and Jeanette M. Wing, *Formal Methods: State of the Art and Future Directions*, CMU Computer Science Technical Report, CMU-CS-96-178, Aug. 1996.

[9] Timothy Colburn, *Methodology of Computer Science* in The Blackwell Guide to the Philosophy of Computing and Information (L. Floridi ed.), Blackwell Publishing, 2004.

[10] ComponentSource - The Definitive Source of Software Components home page, `http://www.componentsource.com/`.

[11] Brad Cox, *No Silver Bullet Revisited*, American Programmer Journal, Nov. 1995.

[12] Brad Cox, ed. *Program Verification: The Very Idea, discussion on Usenet in response to H. Fetzer, Program Verification: The Very Idea, Comm. ACM, Sept. 1988.*

[13] James H. Fetzer, *Program Verification: The Very Idea*, Communications of the ACM 37(9), pp. 1048-1063, Sept. 1988.

[14] Dennis Fowler, *Attack of the Killer Virus*, ACM netWorker, vol. 7, issue 4, Dec. 2003.

[15] Jack Ganssle, *Can Hardware Be Trusted?*, Embedded Systems Programming, vol. 15, no. 2, Feb. 2002.

[16] JAXB (Java Architecture for XML Binding) Standard Implementation Project home page, https://jaxb.dev.java.net/.

[17] Henry E. Kyburg Jr., *Real Logic is Nonmonotonic*, Minds and Machines 11(4), pp. 577-595, Nov. 2001.

[18] Lin, Chu and Liu, *Protocol Verification Using Reachability Analysis: The State Space Explosion Problem and Relief Strategies*, Proc. ACM workshop on Frontiers in computer communications technology, Stowe, Vermont, United States, pp. 126–135, 1987.

[19] Microsoft Research home page, `http://research.microsoft.com/`.

[20] Dejan Milojicic, Alan Messer, James Shau, Guangrui Fu and Alberto Munoz, *Increasing relevance of memory hardware errors: a case for recoverable programming models*, Proc. 9th workshop on ACM SIGOPS European workshop, Kolding, Denmark, pp. 97–102, ACM Press, 2000.

[21] Madanlal Musuvathi, David Park, Andy Chou, Dawson R. Engler and David L. Dill, *CMC: A Pragmatic Approach to Model Checking Real Code*, Proc. Fifth Symposium on Operating Systems Design and Implementation, Dec. 2002.

[22] NASA LaRC Formal Methods Program, `http://shemesh.larc.nasa.gov/fm/index.html`

[23] Nicholas Negroponte, *Bits and Atoms*, Wired Magazine, issue 3.01, Jan. 1995.

[24] Stanford Encyclopedia of Philosophy, `http://plato.stanford.edu/`.

[25] Object Management Group (OMG), *Unified Modeling Language (UML) home page*, `http://www.uml.org/`.

[26] Robert L. Vienneau, *A Review of Formal Methods* in Software Engineering (M. Dorfman and R. Thayer, eds.), Computer Society Press, 1996.

[27] W3C *Extensible Markup Language (XML)*, `http://www.w3.org/XML/`.

[28] The World Wide Web Virtual Library: Formal Methods, `http://vl.fmnet.info/`.

Ontological Approach to Modeling Information Systems

Imad Eldin Ali Abugessaisa
GIS-lab, Department of Computer and Information Science (IDA)
Linköping University, S – 581 83 Linköping, Sweden
g-imaab@ida.liu.se

Abstract

In recent years the use of formal tools in information system modeling and development represents a potential area of research in computer science. In 1967 the term ontology appeared for the first time in computer science literature as S. H. Mealy introduced it as a basic foundation in data modeling.

The main objective of this paper is to discuss the concept of ontology (from a philosophical perspective) as it was used to bridge the gap between philosophy and information systems science, and to investigate ontology types that can be found during ontological investigation and the methods used in the investigation process.

The secondary objective of this paper is to study different design approaches of ontology and ontology development environments that are used to create and edit ontologies.

The paper will discuss ontological engineering as an artifact and approach for modeling information systems. At the end the paper will introduce ontology-driven information systems.

Key words

Ontology, Ontological Commitment, Conceptual Model, Universe of Discourse.

1. Introduction

Aristotle defines ontology as 'the science of being'. This definition can be reformulated as 'the science of being with regards to the aspect of being' [1]. Ontology as a branch of philosophy is the science of what is, of the kinds and structures of the objects, properties and relations in every area of reality. 'Ontology' in this sense is often used in such a way as to be synonymous with 'metaphysics'. In simple terms it seeks the classification of entities[2].

Ontology is descriptive, which means focused on the classification of existing entities [3].

This paper discusses the methods and use of ontology in computer science and information systems modeling.

1.1. Ontology in computer science

Computer science deals with ontologies as modeling or automation tools. Within the artificial intelligence field it is used in knowledge management/knowledge engineering. In ontology-driven information systems, ontologies are used as information integration tool [19]. The classification and reasoning methods that are used by philosophical ontologists can be useful in developing and maintaining ontology in computer science and in information integration in general.

"Ontology is specification of conceptualization," according to the definition given by [4]. Ontology has been defined within its context and it is used for knowledge sharing in AI. The system or program specification gives a detailed description about what should be done by the system, its inputs, processes that will take place in the computer and expected output from the program. Ontology within this framework will provide these specifications and the relationships among the basic entities within the specification. Ontology will also provide knowledge to be shared among different agents (in AI, software agents). There should be general agreement among all parties using the ontology. As a modeling tool, ontology differs from other data modeling tools available in information system development methodologies. Ontology is concerned with the *relationships* among entities rather than with the entities themselves, and with the fact that the semantics of these relationships are applied consistently. In ontology, relationships are defined more or less formally [5] and the semantics of a given relationship is consistently observed. If these relationships are given names that are appropriate to their meanings and human view, the ontology can help the developer to directly understand this relationship.

1.2. Conceptual modeling

Conceptual modeling is a phase of the system development concerned with different system views;

the static view that deals with the static properties of the system such as the entities and relationships[6] and the dynamic view of the system representing how entities are changing their state with respect to time. Some difficulties are associated with conceptual modeling development, such as representation of weak entities in a static model [7]. There are also some problems associated with the modeling of dynamic and temporal aspects [8]. All these problems may affect the information system under development, or the system may fail to reply to user needs. Such difficulties may be caused by the use of traditional conceptual modeling based on descriptive properties of information systems [9].

The roles played by a conceptual model in information system development can be summarized as follows. The representative model should:

*Provide a means of communication between model developers and the end users of the system.

*Increase the system analyst's understanding of the problem domain and all system components.

*Serve as the basis for the design phase.

*Act as documentation of the system requirements.

All the above mentioned roles and good model characteristics should be maintained by modeling tools. Figure 1 shows these features in the system development life cycle.

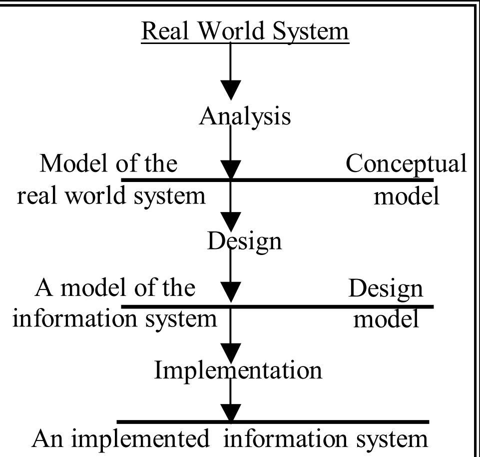


Figure 1: Conceptual modeling.

2. Conceptualization and ontology

System conceptualization is the basic step in the system development life cycle. In both software engineering and system analysis & design, conceptualization refers to the process of specifying system functions and goals and determining system components [10].

After the conceptualization process, a complete vision of the system is developed and all assumptions are validated [11]. Collaborative approach to ontology design [12] adopts a definition of ontology as explicit specification of the abstraction which is used to model the reality into a computer system [13]. Abstraction as a process of generalization will eliminate details and the complexity of reality.

2.1. Universe of discourse (UoD)

Ontologists develop ontology from the domain of knowledge of the real world in which all entities (objects) composing that domain and relationship among them are specified. This domain of knowledge is known as the Universe of Discourse (UoD), which can be defined as 'a complete range of objects, events, attributes, relations, ideas, etc. that are assumed to exist at one occasion'. In a database management system, UoD refers to the part of the world under discussion and it is used to design the schema for [14].

As the UoD maps all relevant aspects of the subject world, the conceptualization (abstraction) must be complete and comprehensive; with respect to the UoD as mentioned before. Ontological commitment is used by agents so they can commit about the specification of the UoD.

2.2. Ontological commitment

Ontological commitment is defined as an agreement and decision by a group of agents (applications) or system users within an application domain to use the terms defined in a given ontology [22]. The advantage of using ontology is that it provides a means for general agreement on the conceptualization specified by all ontology developers and it helps in knowledge exchange. If one of the ontology developers knows something about the entity in the conceptualization it will be easier to represent this to all others in the same UoD. The new knowledge about the ontology in the conceptualization is gained by applying ontological investigation [5]. Ontological commitment was used originally by ontologists in philosophy to get the maximum benefits of sources they have drawn on in their ontological exploration of reality [1].

Ontological commitment provides a means of commitment between agents to exchange data about the UoD. If an agent is consistent in observable actions with the definition of the ontology, it is said to be committed to the ontology. In addition, the common ontology is used to develop ontological commitments for a set of agents within the UoD and all queries and assertions are exchanged among committed agents [1]. Here ontological commitment acts as an agreement to use the shared vocabulary in a coherent and consistent way.

Ontological commitment provides a guarantee of UoD consistency with respect to all queries and assertions using vocabulary defined in the ontology [1].

3. Types of ontology

Different efforts are proposed to classify ontologies. On the basis of use and purposes ontology is either common or a formal ontology [3]. Another classification with respect to the degree of representation of conceptualization and the closeness of the ontology given by [5], fine-grained ontology and coarse ontology.

A coarse ontology consists of a minimal number of axioms and is intended to be shared by users that already agree on a conceptualization of the world. A fine-grained ontology gets closer to the intended meaning of the vocabulary of the knowledge base; while coarse ontology may be characterized as a reference ontology. Coarse ontology users can access these ontologies from time to time for reference purposes (off line) and can support the core system's functionalities [15].

According to the level of generality ontology can be classified into several types: top level, domain and task, application, static, dynamic, intentional and social.

3.1. Top-level ontology

Top level is an ontology which describes general concepts that are independent from a particular problem domain.

3.2. Domain ontology and task ontology

The vocabulary related to a generic domain, task, or activity. This vocabulary can be used to describe the terms introduced in top-level ontology.

3.3. Application ontology

These ontologies use a particular domain and task to describe the concepts that correspond to roles played by domain entities.

3.4. Static ontology

As ontology deals with the existence of entities in the real world, Booch discusses different models of existence for entities which include physical existence, abstraction existence, nonexistence and impossible existence [11].

With such modes of existence of entities, static ontology deals with static aspects of the world which specify the basic attributes of different entities and their corresponding relationships.

3.5. Dynamic ontology

Whereas static ontology describes static aspects of the world, dynamic ontology deals with changing aspects of the world. With dynamic entities, the state of the entity is changed by a process. This process can be a discrete process, in which case it can be modeled by using classical software engineering tools such as Finite State Machine (FSM).

In this special type of ontology each event can be treated as a single time point, and with each event we can have relations such as before and after.

It is also possible to form the concept of causality, which is basically related with time in temporal and dynamic ontology. Simple causality imposes existence constraints on certain actions or events over certain entities.

3.6. Intentional ontology

In order to model different realities and for reasoning purposes intentional ontology should be used. This type includes the 'intentional world', such as intentions, beliefs, interests, goals and choices.

Non Functional Requirements (NFR) [17] can be modeled by using soft goal concepts and is used by intentional ontology to extend it for capturing design rationale.

3.7. Social ontology

Social ontology deals with social settings that are associated with concepts such as actor, position, role, authority, etc, All these concepts are connected with the setting of certain entities such as organizational structure and interdependency. Specialized logics are used to formalize some social concepts.

4. Ontology as engineering artifact

This section will discuss ontology as an engineering artifact that can be compared with the engineering development approach in computer science and knowledge engineering in AI. Ontological engineering deals with the development of ontologies and their use.

Ontological engineering encompasses a set of activities [16], which include philosophical (metaphysic) knowledge representation, formalism, development methodology, etc. In addition to all these activities it can help to give knowledge base a rational design. The definition of essential concepts of the UoD allows the designer to develop a more descriptive knowledge base as well as enabling knowledge accumulation.

4.1. Ontological engineering paradigm

The paradigm of ontological engineering can cover concepts that allow different practitioners from different fields to make the requested conversion between ontological engineering and their descriptions, in order to bridge the gap between these fields.

To meet these objectives an ontology should possess certain qualities [5]: it should be decomposable, extensible, maintainable, modular, interfaceable, tied to the information analyzed being universally understood, translatable and interoperable with a software component or class.

Many similarities between ontological engineering and software engineering allow users from other disciplines to benefit from the computer science discipline. These similarities include software architecture, programming languages, compilers and translators, traditional software engineering, object-oriented analysis and design (OOAD) and design patterns components-based software engineering [8].

4.2. Difficulties with ontological engineering

Many difficulties can be encountered during development of different types of ontology; some of these common difficulties are discussed below.

4.2.1. Development cost and time. Development of formal and automated inference is a difficult and time-consuming task [7]. The reason behind this cost is that getting general ontological commitment among all members of the UoD is very essential.

Two solutions are proposed for ontology developers to take: the first is to develop a small ontology by reaching consensus with a large number of people, where one managing process can be applied over all these ontologies. The second solution is for a certain organization or consortium to build a standard ontology. In the first case certain mapping should be applied and managing process over the small developed ontologies.

4.2.2. User and designer consensus. The aim of building ontology is to support the sharing and reuse of accumulated knowledge by all domain members. This goal faces difficulties because most ontology users do not share the same assumptions. For the users the difficulty is that they cannot easily identify where implicit assumptions are made and where the distinction is due to the ontology itself. All this is considered to be disagreement among the domain members and will be difficult for the designer to deal with in its ontological commitment.

4.2.3. Lack of sound tools. Any engineering field requires sound tools to be used during the engineering process, such as theoretical or empirical techniques that can be used for enhancing the engineering product and to solve certain problems raised during the process.

Theoretical and empirical techniques help the designer to verify and evaluate the process and compare different results. The negative point with ontology development is that ontology engineering does not offer any techniques for evaluation and comparison among different ontologies. Another problem connected with ontological engineering is that there is no possibility to classify problem domains and no support for adequacy and performance measurement of the ontology. It is a challenge for ontology developers to develop empirical research tools and techniques to enable them to evaluate industrial problems.

4.2.4. Information integration. Integration represents an important application area of ontology engineering. Information integration can be obtained when different systems agree upon certain information. If two systems agree upon certain information this means that systems committed to the intended models of the original conceptualization are overlapping [9]. Sometimes two ontologies overlap while their intended models do not.

5. Ontology development

As has been discussed before, ontological engineering is an engineering approach for ontology development; five approaches can be used by an ontologist to create a new ontology or to modify the existing one [13].

5.1. Ontology design approaches

Five approaches to ontological design have been proposed by Holsapple and Joshi]: inspiration, induction, deduction, synthesis, and collaboration [18]. These approaches can be used by ontologists either to develop a new ontology or to modify and enhance an existing one. In addition, all approaches can be used in a single manner or as a combination of two approaches.

5.1.1. Inspirational approach. In this approach ontologists start by specifying the actual needs of the ontology within the UoD. These needs can be specified by individual imagination creating from scratch or by collecting personal views. One of the disadvantages of this approach is that it may lead to narrow ontological commitment.

5.1.2. Inductive approach. By using the inductive approach, an ontologist starts developing ontology using inductive methods and observation. Following these methods for a specific domain (UoD), the collected observations will be examined and analyzed for certain cases; the obtained result will be an ontology that can be simply applied to other cases within the same UoD. The ontologist can measure the quality of the ontology by analyzing the degree of ontological commitment that can be observed from the possibility of applying the ontology obtained from the first case to other cases within the same UoD.

5.1.3. Deductive approach. With this approach the ontologist will work in an opposite way from that followed in inductive methods, starting from adopting general principles and then adapting all these principles and applying them to the UoD in order to construct an ontology for a specific case.
The general principles include both filtering and distilling the general notions that can be customized to a certain case. We can observe ontological commitment in the same way as with inductive methods.

5.1.4. Synthetic approach. Adapting this approach, the ontologist identifies a base set of ontologies. These, along with other selected concepts pertaining to the UoD being investigated, are synthesized to develop a unified ontology. The ontology developed by embracing multiple ontologies gives an opportunity to its adherents to interact in a coherent fashion. Synthesizing the process of getting a unified ontology from a base set of ontologies requires systematic integration of their concepts, elimination of sketches, and reconciliation of different terminologies.

5.1.5. Collaborative approach. This approach for ontology design is based on efforts from different parties reflecting experiences and viewpoints of experts from the UoD and ontologists who intentionally cooperate to produce the ontology. This collaboration will increase chances for high degree of ontological commitment. On the other hand, coordinating the development process needs to be done in a more organized and coordinated manner due to the large number of people involved in this process. The ontologists will apply an iterative process in order to improve it by individual commitment and objections. This self-evaluation process of the collaborative approach makes it the most acceptable one of the five methods.

5.2. Ontology creation

This section is based on Noy and McGuinness's work [20].

5.2.1. Determine the domain and scope of the ontology. In the first step of development the ontologists start defining the domain and scope in which it is intended to develop the ontology. This step can be implemented by specifying the domain that the ontology will cover, the purpose of the ontology, the possible information that should be included in the ontology in order to answer different questions, as well as the users and those who will maintain the ontology.

5.2.2. Reusing existing ontology. In some cases the ontologist will be forced to reuse existing ontologies as a requirement for an existing system that has an existing ontology committed to by the system parties. The reuse of an existing ontology is encouraged by the availability of many ontologies in electronic format that will allow the ontologists to reuse them with the aid of the ontology development environment that is used to edit the same ontology by another ontologies.

5.2.3. Enumeration of terms in the ontology. A list of all terms that will be used within the ontology that users will be able to handle should be specified in this step. All the terms, descriptions and properties should be carefully stated. This is done by developing a comprehensive list of terms, which may contain overlapping concepts. The ontologists should never worry about this overlap in initial listing maintenance, nor the relationship among the terms and the properties of each one. After maintaining the initial list the ontologist will start creating definitions for the concepts within the list.

5.2.4. Definition of classes. At this step of ontology development the ontologist aims to develop a class hierarchy. To perform this step, there are several possible approaches in developing a class hierarchy (Uschold and Gruninger 1996): in the top-down approach the ontologist will start by defining the most general concepts in the UoD and then continue by defining the specialization of each concept, in all steps using the list obtained in section 5.2.3. In a bottom-up approach the ontologist starts by defining the most specific concept within the UoD as a minor class that can be promoted to form the hierarchy of classes, while the group of classes will form a general concept. A combination of these two is a mix of the top-down and bottom-up approaches, where the process starts by defining major concepts first and then generalizes and specifies them appropriately.

5.2.5. Define the properties of classes and slots. The output of the previous step is a hierarchy of classes that represent the concepts and terms within the UoD. At this step the ontologist will start to define the internal structure of concepts. The internal structure of the class will be defined in terms of properties.

The importance of developing such properties is that it will be used as slots attached to classes; several types of object properties can be attached to the class as a slot. All subclasses of a class (x) will inherit the slot of class (X).

5.2.6. Define the facets of the slots. At this step the ontologist starts specifying for each slot its facets. Each slot can have different facets describing the value type, allowed values, the number of values (cardinality), and other features of the values the slot can take. Common facets can be assigned to a slot as slot cardinality, which defines how many values a slot can have. Some systems distinguish only between single cardinality and multiple cardinality. Other types of system allow specification of a minimum and maximum cardinality to describe the number of slot values more precisely. Types of values that can be held by a slot are string, number, or Boolean.

5.2.7. Creation of instances. This is the last step the ontologist will apply to get the required ontology for a specified UoD. The step includes creation of individual instances of each class in the hierarchy. The definition of the instances requires first choosing a class then creating an individual instance of that class, and filling in the slot values.

5.3. Ontology development environment

An ontologist will need tools to create a new ontology or edit an existing one. A general environment to facilitate the development and sharing of ontologies is provided and it can be used to model and share the knowledge domain. All available environments provide ontology developers with the basic development tasks of browsing, creating, maintaining, sharing, and using ontologies.

5.3.1. Knowledge interchange format. This is a formal language for the interchange of knowledge among disparate computer programs [21]. It acts as a mediator in the translation of languages. Similar to any programming language, it includes both syntax and semantics specification. KIF provides the developer with variety of logical axioms to be used for encoding logical information. KIF is also used for encoding knowledge about knowledge and it can help to create and describe procedures. KIF has a semantics similar to first order logic.

5.3.2. Chimaera. A Web-based software system that supports the ontology developer in creating and maintaining distributed ontologies over the web. An ontology developer can perform two basic functions with Chimaera [20], merging multiple ontologies together and diagnosing individual or multiple ontologies. Chimaera also helps to maintain such tasks as loading knowledge bases in different formats, reorganizing taxonomies, resolving name conflicts, browsing ontologies and editing terms. This software is freely available with full documentation and online help.

5.3.3. Ontolingua. Another Web-based ontology development environment that could be used to browse, create, edit, modify, and use ontologies. The server supports over 150 active users and it has been available over the Web free from Stanford KSL Network Services [23].

6. Ontology-driven information systems

The purpose of using ontology is either to share and reuse information or to support specifications, i.e. entities and relationships between them. The translation of ontology into an active information system component leads to Ontology-Driven Information Systems (ODIS). When an explicit ontology plays a central role in the system life cycle, the ontology drives all aspects and components of the system. Ontologies can be used at development time or at run time. In ODIS the ontology is called application ontology [19].

6.1. Ontology-Driven Geographic Information Systems

ODGIS are used to solve problems associated with availability of data to the users and the documentation of metadata. During development of ODGIS, ontologies are translated into software components, which are classes with knowledge embedded, and they can be reused for developing new GIS applications. Problems handled by ODGIS are data availability and metadata-associated problems, the possibility of performing queries based on semantic values, the availability of information at different levels of detail, and dynamic access to information that otherwise would be difficult to obtain. All mentioned problems may relate to interoperability or semantic granularity of the GIS data [19].

7. Conclusion

This paper investigates the ontology approach to modeling information, as ontology can act as an information integrator. The paper discusses the basic concepts related to philosophical ontology as well as the use of the concept in different computer science fields. The fundamental issues of the conceptualization and the conceptual model have been discussed because that is the phase of system development in which the ontology can help the developer to develop a system conceptual model by using ontology principles. Moreover, some basic approaches for ontology design are discussed. The problems of ontology development are discussed with the possible solutions that can be followed. Different tools that can be used by ontology developer which act as ontology development environment (ODE) are introduced.

The paper discusses two cases in which ontology is used: ontology-driven information systems (ODIS) and ontology-driven geographical information systems (ODGIS).

8. Acknowledgments

I am grateful to Gordana Dodig-Crnkovic for her generosity with her time and insights. I would like to thank the three reviewers as well as my colleagues Rikard Land for his comments to modify and extend the paper.

9. References

[1] G. L. Zúñiga, "Ontology: Its transformation From Philosophy to Information Systems", Proceedings of the International Conference on Formal Ontology in Information Systems, Ogunquit, Maine, USA , 2001, pp. 187-197.

[2] J. Alexander, M. Freiling, S. Shulman, J. Staley, S. Rehfuss and S. Messick, "Knowledge Level Engineering: Ontological Analysis", Proceedings of the 5th National Conference on Artificial Intelligence, Published by AAAI Press, Menlo Park, California USA , 1986.

[3] S. Bary "Towards an Ontology of Common Sense", in Jaakko Hintikka and Klaus Puhl (eds.), The British Tradition in Twentieth-Century Philosophy , Vienna: Hölder-Pichler-Tempsky, 1995, 300-309.

[4] T. R. Gruber, "A Translation Approach To Portable Ontologies", Journal of Knowledge Acquisition, Vol. 5(2), Academic Press Ltd. London, UK, UK 1993, pp. 199 - 220.

[5] N. Guarino, "Formal Ontology in Information Systems", Proceedings of FOIS'98, Trento, Italy, 6-8 June 1998. Amsterdam, IOS Press, pp. 3 - 15.

[6] P. Pin and S. Chen, "The Entity-Relationship Model : Towards A Unified View of Data", ACM Transactions on Database systems, ACM Press New York, NY, USA , 1976 , pp. 9 - 36.

[7] Y. Wand, V. C. Storey, R. Weber, "An Ontological Analysis of The Relationship Construct in Conceptual Modeling", ACM Transactions on Database Systems (TODS) Volume 24, Issue 4, ACM Press New York, NY, USA, pp. 494 - 528 .

[8] H. Gregersen and C. S. Jensen, "On the Ontological Expressiveness of Temporal Extensions To The Entity-Relationship Model", Proceedings of the 9th International Conference on Entity-Relationship Approach (ER'90), ER Institute Lausanne, Switzerland,1999, pp.249 - 264.

[9] Holm, P, O*n The Design And Usage of Information Technology And The Structuring of Communication At Work*, Stockholm University press, Stockholm, 1996.

[10] P. Spyns, R. Meersmann, M. Jarrar, "Data Modeling Versus Ontology Engineering", ACM SIGMOD Record , ACM Press New York, NY, USA , 2002 , pp. 12 – 17.

[11] G. Booch, *Object-Oriented Analysis and Design with Applications*, Addison-Wesley 2nd edition, Boston, 1993.

[12] C. W. Holsapple, K. D. Joshi, "A collaborative Approach To Ontology Design", Communications of the ACM, 45(2) ACM Press New York, NY, USA , pp. 42 - 47.

[13] Bernhardsen, T, *Geographic Information Systems: An Introduction*, 3rd ed, John Wiley & Sons Inc., New York, 2002.

[14] Date, C. J., *An Introduction to Database Systems*, Addison-Wesley, Boston , 1990.

[15] A. Stutt., "Knowledge Engineering Ontologies, Constructivist Epistemology, Computer Rhetoric: A

Trivium for the Knowledge Age", Knowledge Media Institute, The Open University, UK, 1997.
[16] V. Devedzić, "Understanding Ontological Engineering", Communications of the ACM Volume 45, Issue 4, ACM Press New York, NY, USA, 2002, pp. 136 – 144.
[17] I.Sommerville, *Requirements Engineering: A Good Practice Guide*, John Wiley & Sons , Boston , 1997.
[18] C.W. Holsapple and K.D. Joshi , "A collaborative Approach to Ontology Design", Communications of the ACM Volume 45 , Issue 2 , ACM Press New York, NY, USA , 2002, pp. 42 – 47.
[19] F.T. Fonseca and M. J. Egenhofer, "Ontology-Driven Geographic Information Systems", Proceedings of the seventh ACM international symposium on Advances in geographic information systems, ACM Press New York, NY, USA , 1999, pp.14 – 19.
[20] N. F. Noy and D. L. McGuinness, "Ontology Development 101: A Guide to Creating Your First Ontology", Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880, March 2001.

Web Resources :

[21] Knowledge Interchange Format. Available at http://logic.stanford.edu/kif/kif.html (Accessed February 25, 2004).
[22] Chimaera at Knowledge Systems Laboratory, Stanford University. Available at http://www.ksl.stanford.edu/software/chimaera/ (Accessed February 23, 2004).
[23] Ontolingua. Available at http://www.ksl.stanford.edu/software/ontolingua/ (Accessed February 23, 2004).

Complexity as a Matter of Information

Alexander Backlund
School of Humanities and Informatics
University of Skövde
alexander@ida.his.se

Abstract

In this paper, some work on the concept of complexity is reviewed, and it is shown how many issues concerning complexity in general, and concerning organisations and information systems in particular, can be understood in terms of information and information processing, e.g. lack of information, abundance of information, and information loss. Complexity is closely related to our cognitive abilities and limitations, and study of complexity therefore encompasses study of cognitive science.

1. Introduction

Issues regarding complexity have been considered to be of great importance, and have over time attracted attention in various disciplines, not least in systems science. Klir [1] writes:

> "Systems complexity is primarily studied for the purpose of developing sound methods by which systems that are incomprehensible or unmanageable can be simplified to an acceptable level of complexity. Problems of systems simplification are perhaps the most important of all systems problems. Gerald Weinberg [(1975)] goes even further when he defines systems science as a science of simplification and argues well the importance of methods of simplification […]" (p. 135)

And Börje Langefors [2] identifies six fundamental problems of systems development. Complexity is one of them.

This work aims to discuss complexity in terms of information, and to demonstrate that complexity in many instances, if not all, can be seen or understood as a matter of information and information processing, e.g. lack of information, abundance of information, and even information loss.

Now, there is no general agreement as to what information and complexity really are or should be defined as. Different viewpoints will be adopted throughout this work, though basically, the infological view on information and information systems advocated by Langefors (e.g. [2]) will be adopted. This is a view according to which information does not exist without humans (or, one could imagine, some other sentient being capable of interpreting data) – and therefore all information systems include humans but not necessarily computers and suchlike tools, and every organisation can be seen as an information system, since every organisation consists of humans directly or indirectly communicating, i.e. exchanging information, with each other. Could a person really be a part of an organisation if he/she never in any way communicated with any other member of that organisation? No, that is not conceivable. Langefors' [2] infological equation states that $I=i(D, S, t)$, where i is an interpretation process that produces the information I from the data D and the pre-knowledge S during the time period t.

> "The infological equation makes it clear that the people who are to interpret the data are included in the information system. The information entites in the systems are not data, since data are not information. It follows that data systems, by themselves, are not information systems. A data system must be combined with the organisation, or part of it, before an information system emerges." [2] (p. 144-145)

This is an appealing and often useful viewpoint. However, it is difficult to use when quantifying information, and sometimes *information* will instead be used as a synonym to *data*.

(There are of course other ways to define information, like that of Stafford Beer [13], who defines information as "*That which CHANGES us*." He says that we know that we have been informed only because our sate has changed. Similar definitions were given by Bateson and MacKay (see [14]). While there is arguably a point in these definitions, they still seem too broad. Let us say a man is having a walk. Suddenly, a roofing-tile falls down, hits him on the head, immediately killing him. Certainly, his state has changed, and the event might change the state of others too and generate information that they receive. However, it does not seem reasonable to say that that unfortunate man has received any information.)

Also, the fundamental view on complexity adopted here is that of Backlund [3], who sees complexity as a basically subjective property (though it could also be regarded as intersubjective) and defines it thus:

> "Since complexity is something perceived by an observer, the complexity of the system being observed is, one could say, a measure of the effort, or rather the perceived effort, that is required to understand and cope with the system." [3] (p. 31)

This means, of course, that the complexity of a system is not fixed. The complexity does not depend solely on its properties and can change. E.g. when we learn more (and thus have acquired and/or incorporated more information into our existing pre-knowledge), the same thing, be it an object or a system, can seem less complex than it did at first. However, complexity is not totally independent of the properties of the object or system that is observed either. The complexity of something is not determined arbitrarily, though, but it depends on the properties of the object and the strains that those properties put on our cognitive abilities when we try to understand and cope with it. (And however different we may be, our cognitive abilities are yet fairly alike – as a rule.) Flood and Carson [4] see complexity as a property of things and the people who observe them.

For a definition of what a system is, see [15]. However, basically, a system is something which consists of two or more elements which have a direct or indirect connection to all other elements in the system.

1.1. Stress

Miller's living systems theory [5] encompasses a number of different kinds of systems: cells, organs, organisms, groups, organisations, communities, societies, and supranational systems. These systems need to have the 20 critical subsystems or to have the processes of these performed carried out by some other system (e.g. a suprasystem or a system in its environment). The subsystems process matter-energy and/or information (which is defined as negentropy – "It was noted by Wiener and by Shannon that the statistical measure for the negative of entropy is the same as that for information, which Schrödinger has called 'negentropy.'" (p. 13)). There are a number different types of stresses that might occur in living systems. Now, what do we mean by stress?

> "There is a *range of stability* for each of numerous variables in all living systems. It is that range within which the rate of correction of deviations is minimal or zero, and beyond which correction occurs. An input or output of either matter-energy or information which, by lack or excess of some characteristic, forces the variables beyond the range of stability, constitutes *stress* and produces a *strain* (or strains) within the system. Input lack and output excess both produce the same strain – diminished amounts in the system. Input excess and output lack both produce the opposite strain – increased amounts. Strains may or may not be capable of being reduced, depending upon their intensity and the resources of the system. […]" [5] (p. 34)

There are different kinds of information stress:

> "Systems also undergo *information stresses*, including: (a) information input lack or underload, resulting form a dearth of information in the environment or from improper function of the external sense organs or input transducers; (b) injection of noise into the system, which has an effect of information cutoff, much like the previous stress; and (c) information input excess or overload. Informational stresses may involve changes in the rate of information input or in its meaning." [5] (p. 35)

Now, let us remember what has been said before about the cognitive aspects of complexity and consider what Rescher [6] has to say about it:

> "In general […] cognitive difficulty reflects […] complexity. As a rule, an item's complexity is indicated by the extent to which we encounter difficulty in coming to adequate cognitive terms with it. By and large, the amount of effort that must be expended in describing and understanding the make-up and workings of a system is our best practical indicator [of] complexity, and its inverse is our best practical indicator of simplicity." (p. 17)

Now, we can roughly equate Miller's [5] eleven critical information processing subsystems in a human with our cognitive system. (There are ten critical subsystems that process information and one, boundary, that processes both mater-energy and information. I have left the reproducer out of consideration.) And we can consider complexity from the following aspects:

- Lack of information or uncertainty (in information theory information is often defined as reduction of uncertainty, which makes uncertainty a matter of lack of information): “information input lack or underload” and “injection of noise into the system”.

- Abundance of information: information input excess or overload.
- Information loss (which has no direct relation to the different kinds of stresses discussed above but has a connection to our cognitive limitations – as well as strengths – as will be seen below).

2. Complexity as an aspect of information processing

Complexity is determined by our cognitive processes and what causes strain on them. “A thing is complex when it surpasses human cognitive limitations” [2] (p. 87).

2.1. Lack of information or uncertainty

It is well-known that vast amounts of information can contribute to complexity (see below), but it is also important to acknowledge that complexity can arise out of ignorance, lack of information or uncertainty regarding the properties and relations of something (e.g. an object, a system, a situation). Complexity can also be related to change, which might involve both uncertainty and vast amounts of information. Flood and Carson [4], interpreting Vemuri (1978, in [4]), write:

> “1. Complex situations are often partly of wholly unobservable, that is, measurement is noisy or unachievable (e.g., any attempt may destroy the integrity of the system).”
>
> “2. It is difficult to establish laws from theory in complex situations as there are often not enough data, or the data are unreliable so that only probabilistic laws may be achievable.”
>
> “3. Complex situations are often soft and incorporate value systems that are abundant, different, and extremely difficult to observe or measure. They may at best be represented using nominal and interval scales.”
>
> “[---]”
>
> “4. Complex situations are ‘open’ and thus evolve over time.” (p. 120)

In a small series of interviews with systems analysts, Backlund [7] found that one aspect which made information systems seem complex to them was the degree of ambiguity concerning rules and conditions.

Weaver’s (1948, in [4]) three ranges of complexity can also be discussed in this context. The three ranges of complexity identified by Weaver are organized simplicity, organized complexity, and disorganized complexity. Flood and Carson [4] would also like to add something they call the people range. When the number of significant factors are small and the number of insignificant ones is large, then organized simplicity occurs. A situation might seem complex at first, but when all insignificant factors are eliminated the underlying simplicity of the situation is revealed. According to Klir (1985b, in [4]), that kind of discovery is typical of the 17-19th century science. Of course, here, what causes complexity is not primarily lack of information but rather and *abundance* of irrelevant information. However, this abundance of information would not cause complexity unless there was a *lack* of information and an uncertainty regarding *which* information is relevant.

For disorganized complexity to occur there must be many variables the behaviour of which is highly random. An example of this is the behaviour of gas molecules (Klir, 1985b, in [4]). Organized simplicity can be mathematically dealt with analytically, and disorganized complexity statistically. What causes complexity in this case could also be said to be uncertainty or lack of information, because there is a fundamental uncertainty as to how the parts of the system will behave (while relatively accurate predictions can be made on the system as a whole).

(“Typical of organized complexity is the richness that must not be oversimplified, but equally cannot be dealt with by techniques that work effectively on a large degree of randomness.” [4] (p. 35)

When it comes to the people range, “What characterizes this dimension is plurality. Each situation may be appreciated in different ways by different people.” [4] (p. 35) This can be regarded in many ways. However, here it is a matter of fundamental uncertainty not only about how to relate facts to each other but also about facts themselves (depending on, among other things, experience, knowledge, and *Weltanschauung*).

2.2. Abundance of information

Abundance of information often causes complexity, and we will look into a few aspects of this. E.g. Backlund [2] found in the previously mentioned study that one aspect which seemed to create complexity in information systems is a diversity and multitude of rules. Other aspects that required handling a lot of information were mentioned.

We have already considered the role of lack of information in relation to Weaver's ranges of complexity. However, we can also see that complexity in these cases also arise from an abundance of information. In the case of organized simplicity, it seems reasonable to assume that a lot of the complexity comes from the fact that we have a lot of information about many factors which obscures the relevant information about the relevant factors. When dealing with disorganized complexity, there is a high degree of randomness, which means that there is potentially a vast amount of information about the behaviour of the individual parts. However, seen as a whole system, the amount of information and the complexity drastically decrease when statistical laws are applied to it. In the case of organized complexity, "the richness that must not be oversimplified" [4] (p. 35) means that there is a lot of information which has to be regarded.

Langefors [2] says: "Complexity is the property of being a thing that can only be perceived piecewise" (p. 70), and as is pointed out by e.g. Backlund [3], because of the well-known limitations of our short-term memory (7±2 chunks, according to the classical paper by Miller [8]), it is tempting to say that something is complex when it cannot be represented in 7±2 chunks. However, it is also pointed out that a telephone directory cannot by far be chunked into 7±2 chunks, and it is still not considered particularly complex. It would seem then that the sheer amount of information is not enough to make something complex. Perhaps this can be attributed to the *manner* in which the information is processed (which might also depend on the way the information is organised).

Still, Ashby [9] (p. 1), who sees "a system's complexity [as] purely *relative to a given observer*", i.e. it is "something in the eye of the beholder", suggests that we should measure complexity "*by the quantity of information required to describe the vital system.*" (This is, of course, a most subjective measure. The length of the description required differs from one observer to another, depending on interests and knowledge, and also on the models used to describe it.)

We will briefly return to this definition shortly. First, however, we should consider the characterisations of complexity by Brewer [10] and Yates [11]. Brewer [10] (p. 7) writes: "As a model's elements become increasingly interconnected, it becomes increasingly complex." And Yates [11] (p. r201) says:

> "complexity usually arises whenever one or more of the following five attributes are found: 1) significant interactions; 2) high number (of parts, degrees of freedom, or interactions); 3) nonlinearity; 4) broken symmetry […]; and 5) nonholonomic constraints".

Now, generally, it seems reasonable to assume that these properties in a system create increased complexity and that more information is needed to describe such systems. Is this a rule with no exception? Do e.g. a large number of elements or relations always make a system complex? Backlund [3] gives an example of a system that does not seem to be very complex (though the reader might consider the fact that the complexity of the system seems greater represented symbolically than graphically):

> "Consider the following system, *S*, consisting of the set M and a set of relations on *M*, *R*:

$$M=\{a_1,..,a_n,b_1,..b_n\}$$

$$R=\{(x,y)\,|\,(x=a_1\wedge y=b_1)\vee(x=b_1\wedge y=a_1)\vee(x=a_n\wedge y=b_n)\vee(x=b_n\wedge y=a_n)\vee$$
$$\vee(x=a_m\wedge y=a_{m+1})\vee(x=a_{m+1}\wedge y=a_m)\vee(x=b_m\wedge y=b_{m+1})\vee(x=b_{m+1}\wedge y=b_m),1\leq m\leq n-1\}$$

> The system is not very complex for n=3, and not for n=6 (as is illustrated in Fig. 1), and it will not be much more complex for n=10 000 either. (As well as by just looking at the system, by considering Ashby's [9] definition of complexity […], where the complexity of a system is measured by the amount of information needed to describe it, we can see that the complexity does not increase significantly.)" (p. 33)

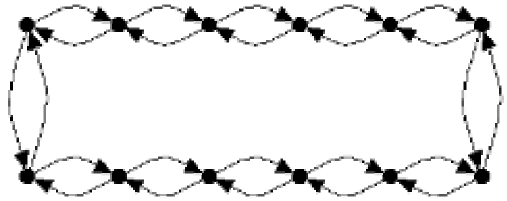


Fig. 1: A system. At the top: $a_1,...,a_6$. At the bottom: $b_1,...,b_6$. (From [3], p. 33)

The information needed to describe the system does increase slightly, but not considerably, and intuitively, the complexity remains more or less the same – *in this case*.

Understanding can be defined in the following manner:

> "'Verstehen … [ist] das Erfassen von Zusammenhängen' (Brockhaus Enzyklopädie, 1994, p. 272). A similar definition in English would be: Understanding is awareness of connections. To understand an organisation would thus be to be aware of the connections between it and other objects in its environment and between its parts, and the nature of those connections. A connection is a relation or an association." [12] (p. 8)

The more elements, relations, and interactions, the more there is to know about the system, and the more information is needed to understand the system, and the more information that is needed, the greater the cognitive effort required to understand and cope with the system. Complexity and difficulty to gain an understanding are thus interrelated, and both are, in many cases, increased by a large amount of information that needs to be properly gathered, processed and grasped.

2.3. Loss of information: a measure of complexity

Abundance of information often causes complexity, and we will look into a few aspects of this. E.g. Backlund [7] found in the aforementioned study that one aspect which seemed to create complexity in information systems is a diversity and multitude of rules. Other aspects that required handling a lot of information were mentioned.

Backlund [3] discusses at some length the correlation between how much information (in the infological sense) that "disappears" in an organisation/information system (remember, every organisation is an information system, and normally every information system serves an organisation) and the complexity of that organisation/information system. However, basically, if the information system (and/or its processes) is complex, it is likely that information is lost, changed, or compiled so that the amount of information that reaches the decider is only a fraction of what was originally received by the information system or created within it.

"Let $a_1, a_2, \ldots, a_n$ be elements of the information system in which information is originally created or received and forwarded from outside the system. Let the total amount of information originally created by or received by $a_1, a_2, \ldots, a_n$ be denoted by I_a. Let the total amount of information received by the highest echelon of the decider subsystem in the organisation be denoted I_d. The information common to I_a and I_d is denoted I_c. Let $g(x)$ denote a function returning some measure of the amount of information in x. E is the efficiency of the information system. The complexity of the information system is denoted C_{is}." [3] (p. 38) Then

$$E = \frac{g(I_c)}{g(I_a)}$$

$$C_{is} = 1 - E$$ [3]

The main difficulty here is that there is no way to know if and to what extent the information received by person A is the same as person B wished to convey, not even if they use the same words to express their ideas. Also, it is difficult to quantify.

3. Summary

Complexity can be viewed in many ways. However it is viewed, though, it seems that complexity can always be seen, directly or indirectly, as a matter of information and information processing, whether it be too much information or too little; both cases seem to generate complexity (though not always). And perhaps this should not be surprising, because there is a clear link between complexity and the limitations of our cognitive abilities; our cognitive subsystems process information in various ways to allow us to perceive and handle the world around us. It also seems that the nature of our cognitive processes and information processing in general makes information "disappear" on its way up the echelons (levels of command in the decider subsystem) in an information system. This is necessary, since the upper levels would otherwise suffer from information overload. However, there might be a connection between how much information that disappears (i.e. does not reach the top echelon of the decider at all or is transformed in different ways) and how complex the information system is (see ch. 2.3 and [3]).

4. References

[1] G.J. Klir, "Complexity: Some General Observations", *Systems Research*, Vol. 2, No. 2, pp. 131-140, 1985.

[2] B. Langefors, *Essays on Infology*, Studentlitteratur, Lund, 1995.

[3] A. Backlund, "The concept of complexity in organizations and information systems", *Kybernetes*, Vol. 31, No. 1, pp. 30-43, 2002.

[4] R.L. Flood and E.R. Carson, *Dealing with Complexity. An Introduction to the Theory and Application of Systems Science. Second Edition*, Plenum Press, New York, 1993.

[5] J.G. Miller, *Living Systems*, University Press of Colorado, Colorado, 1995.

[6] N. Rescher, *Complexity: A Philosophical Overview*, Transaction Publishers, New Brunswick, 1998.

[7] A. Backlund, "What Makes Information Systems Complex in the Eyes of Systems Analysts?" In: J.K. Allen, and J. Wilby (eds.), *Proceedings of the 46th Annual Conference of the International Society for the Systems*

Sciences, August 2-6, 2002, Shanghai, P.R. China. CD-ROM.

[8] G.A. Miller, "The magical number seven, plus or minus two: some limits on our capacity for processing information", *The Psychological Review*, Vol. 63, No. 2, pp. 81-97, 1956.

[9] W.R. Ashby, "Some Peculiarities of Complex Systems", *Cybernetic Medicine*, Vol. 9, No. 2, pp. 1-7, 1973.

[10] G.D. Brewer, *Analysis of Complex Systems: an Experiment and its Implications for Policymaking. Report no. P-4951*, The Rand Corporation, Santa Monica, California, 1973.

[11] F.E. Yates, "Complexity and the limits of knowledge", *The American Journal of Physiology*, Vol. 4, pp. r201-r204, 1978.

[12] A. Backlund, "How to apply living systems theory to the analysis of organisations". In: J.K. Allen and J. Wilby (eds.), *Proceedings of The World Congress of the Systems Sciences and ISSS 2000 International Society for the Systems Sciences 44th Annual Meeting*, Toronto, Canada, July 16-22, 2000. CD-ROM.

[13] S. Beer, *The Heart of Enterprise*, 2nd ed., John Wiley & Sons, Chichester, 1979.

[14] L. Floridi, "Information". In L. Floridi (ed.), *The Blackwell Guide to the Philosophy of Computing and Information*, Blackwell Publishing, Oxford, pp. 40-61, 2004.

[15] A, Backlund, "The definition of *system*", *Kybernetes*, Vol. 29, No. 4, pp. 444-451, 2000.

Ethics of the hyperreal

Thomas Larsson
Department of Computer Science
Mälardalen University
Sweden

***Abstract*— In this essay, it is argued that virtual reality constitutes an extremely powerful learning environment raising serious ethical issues. Technology advancements make the created illusions and experiences increasingly realistic, bridging the gap of the noumenal and phenomenal, the real and hyperreal. Applications, fooling the mind, can give rise to an uncontrolled moral looping affecting our real selves. Therefore, it is of great importance that the ethical implications of the usage and creation of virtual reality applications are discussed and better understood.**

***Index Terms*— ethics, virtual reality, simulation, video games, education, human rights, crimes, violence, harm.**

I. Introduction

Virtual reality is about creating the most convincing illusion possible of actually experiencing, through the human senses, an alternative non-existing world — a virtual reality. As early as 1965, computer graphics pioneer Ivan Sutherland wrote:

> The screen is a window through which one sees a virtual world. The challenge is to make that world look real, act real, sound real, feel real. [20]

We have seen a tremendous technology development since then. Virtual reality has found uses in scientific visualization, medical simulation, virtual prototyping, architecture, teleoperation of robots, flight simulation, communication, recreation and entertainment. It is true that many of the applications created in these areas have been found beneficial and uncontroversial.

Nevertheless, we must not forget that virtual reality is about fooling the minds of the users by giving them a profound sensation of actually being present in an artificial world. This is the very goal of virtual reality technology and this is what makes virtual reality such a powerful medium. In Figure 1, the active and realistic nature of the virtual reality medium is illustrated. The most important factors in achieving realism are immersion, interactivity, and sensory feedback. This is in stark contrast to, for example, a traditional theater where the viewers are passive watchers. The viewers can of course dream themselves into the plot, however, this is a much weaker form of realistic experience.

It is now technically possible to build a virtual reality system capable of creating highly realistic interactive experiences and illusions to fulfill many of a user's wishes. Several human activities, for example, communication, art, politics, violence, and sex, have found a new home in virtual reality [19]. The moral consequences and social issues this gives rise to must be considered and understood. Many people worry about that some users will live out their fantasies in an unrestricted way by creating advanced alter-egos, as it may seem, free from moral obligations and consequences. To a certain degree, this is already possible in the prevalent role playing games of today, in which thousands of users interact and form on-line communities[1].

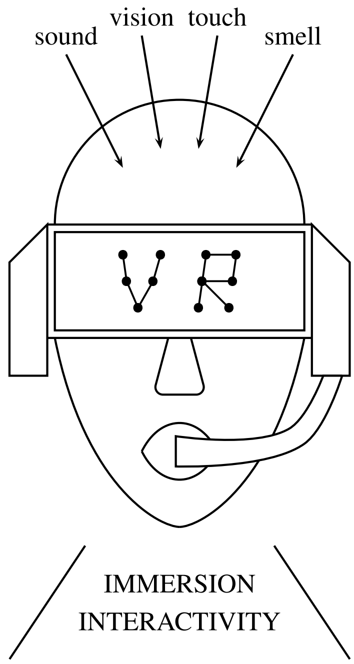


Fig. 1. Traits of the virtual reality medium. The user perceives an artificial world by means of immersion, interactive actions, and multi-sensory feedback.

The role of immersion in virtual reality applications is clearly a reason for concern. For example, constantly improved simulation technology enables increasingly immersive computer games which can cause unwanted effects on society [9]. Various virtual reality platforms are already available to video game consumers. As the immersive experience gets more and more realistic, young impressionable minds might find it harder to distinguish reality from fiction. As the French philosopher Jean Baudrillard points out, the situation is already complicated. He claims that we already live in a society where TV, newspapers, and computer images are more real to us than the physical reality in which we are embodied. We live in a world with made-up models of reality; that is,

Manuscript created May 19, 2004

[1]In multiple-user simulations it is possible to act unethically by directly offending or hurting other humans. The discussion herein, however, is mainly focused on single-user virtual reality experiences and its ethical implications.

a hyperreality built on simulations (images) and simulacra (signs). Since realistic simulation involves more than just pretending, the difference between true and false, the real and the imaginary is threatened, and the truth principle is challenged. As Baudrillard says:

> To dissimulate is to feign not to have what one has. To simulate is to feign to have what one hasn't. One implies a presence, the other an absence. But the matter is more complicated, since to simulate is not simply to feign: 'Someone who feigns an illness can simply go to bed and pretend he is ill. Someone who simulates an illness produces in himself some of the symptoms'. [4]

Leaving the most provocative and far-reaching conclusions drawn by Baudrillard, as he elaborate his thesis, aside, such as "illusion is no longer possible", we still must face the point emphasized here: virtual reality simulations threaten us by its potential to literally confuse our mind, personality, and worldview. What if the immersion gives rise to a blurred world view of mixed realities with contradicting goals. What would the consequences be? According to Schroeder [18], it is here the issue of video game violence — the far most prevalent form of simulation of violent actions — enter a new realm where the awareness of consequences to action collapses as well as the need for ethics:

> There is no ethics of the hyperreal. The potential problem with video-game culture and the simulation theories that describe it is not, then, an actual leakage of the playspace, but an electronically induced amnesia. Video games do not teach the wrong ethics, they teach that ethics are superfluous: only the game counts, and the game can be started over and over again. This looping recursive world is fine if we remember that its existence is confined to a playspace. But in virtual reality we may be tempted to forget. And if immersive media, as increasingly 'real' environments, teach us to forget the parameters of play, they do it in conjunction with simulation models that theorize the hyperreal. The problem is to maintain a 'structuralist' critique of a medium that increasingly threatens post-structuralist collapse, for if the world itself becomes a playspace, then accountability and ethics drop out. There are no more consequences, except the need to push *restart*.

Since virtual reality is about fooling the mind, people legitimately fear that demoralized applications might lead to severe negative effects on society. For example, laws of the real world might be undermined and suppressed if virtual worlds promote or encourage alternative social rules and ethics, contradicting their real world counterparts. Community values might be eroded and compromised. Therefore, virtual reality prospectives give rise to new important political and ethical questions.

Science fiction writers and storytellers often imagine or try to foresee a future in which the real and the virtual worlds are intertwined and deeply dependent upon in each other, sometimes in bright utopian settings, but in most cases with dark dystopian undertones. For example, in the first cyberpunk novel, Neuromancer, William Gibson envisions a society deeply dependent on an artificial-intelligence dominated virtual universe. A global network of interconnected computers — cyberspace — provides "consensual hallucination" and sharing of experiences that looks and feels like a physical space. In this environment, an ex-hacker is recruited to fight an extremely powerful artificial intelligence, had he not been deceived [12].

Another example can be found in the movie The lawnmower man (1992), mostly known for its use of at the time cutting edge computer graphics effects. In this case, a badly treated simple-minded gardener is turned into a genius by a learning experiment termed "intelligence enhancement". A combination of virtual reality technology and drugs increase the subject's IQ, perception, and strength significantly, but eventually, the experiment goes out of hand and ends up in a technology empowered rampage. Other examples of science fiction stories that relate to virtual reality includes the Holodeck in Star Trek, and the more recent Matrix movie, in which the experienced world is a result of a perfect simulation indistinguishable from the real world.

As the virtual reality technology develops, scenarios resembling science fiction, like the mentioned examples, may become possible to experience, alter and live out. For example, the era of cyborgisation has already been opened, which involves the merging of flesh and machines. This area also includes severe ethical implications [21].

One of the far reaching ideas that also has begun to be explored is the retinal display invented 1991. In this case, the image is formed directly on the retina by shining a modulated laser in the eye. The advantages of this approach is the very high resolution images that can be created and the precise eye tracking it makes possible [15]. This technology might also make high quality wearable virtual reality devices a feasible alternative from which we can expect many novel applications.

In what follows, some of the ethical implications of human activities and behavior made possible by virtual reality are discussed. In particular, the educational power inherent in the medium will be elaborated and used as a vantage point for the ethical discussion.

II. Educational power

Virtual reality applications are sometimes praised for their educational potential. Inherent in all virtual realities is a pedagogy, or a learning environment, based on attention, repetition, and reinforcement. On the positive side, applications can, for example, help the users to develop various problem solving and communication skills, and they can even be of therapeutic benefit in many medical contexts.

On the other hand, the virtual reality technology may be misused in several ways. Hurtful and demoralizing applications may become commercially successful and prevalent, particularly among children and adolescent, much in the same way as has already happened with some highly violent computer games such as Grand Theft Auto 3 and its sequel Grand Theft Auto, Vice City. In these games, the user's mission involves crimes like stealing, drug delivering, and killing, and violence on law enforcement officials and women is actually rewarded. There are many other examples of hyperviolent computer games on the market [6]. What lessons are the users learning from them? How does it affect our society?

Virtual reality applications constitute an even greater threat by offering a more powerful educational paradigm. It should be clear that through the simulations, the language of ethics, or the lack thereof, is inevitably spoken to the attentive and focused users. What are the users learning and with what consequences? If we consider the development of computer games, and the commercial success stories of highly violent titles with increasingly realistic graphics, the question is certainly justified. The number of available studies on the effects of highly violent video games gives us strong reasons to believe that these games cause aggression, violent attitudes and behavior, and decrease prosocial behavior. They also seem to desensitize children to violence and create a climate of fear [2], [7], [11].

A. The social cognitive theory

The social cognitive theory by Albert Bandura was officially published in 1986 [3]. This theory stems from the earlier social learning theory in which learning rests on the principles of reinforcement, punishment, extinction, and imitation. These are social learning factors determining the type, frequency, and targets of aggressive behavior.

The social cognitive theory extends this view by recognizing that behavior is largely based on cognitive processes, where behavior results from a recursive interaction of personal factors, the environment, and behavior. This suggests that a person's reality is formed through a constant flow of cognitions, interactions, and feedback. To understand human behavior, we need to understand the involved cognitive processes as well.

Although the comprehensiveness of the social cognitive theory make it difficult to use in specific applications, the theory strongly suggests that most behavior is based on activities in our mind and learning. This has clear implications on virtual reality, since the cognitive processes and the active learning environment it creates on the user are very real, and will therefore play an important part among the driving forces of our actions. Interestingly, experimental studies on the effects of violent video games lend support to this theory, since in general, participants become more aggressive after observing violent imagery, and it is not uncommon that they start imitating the observed aggressive characters [9], [10].

Role models play a significant part in our social learning. Vicarious reinforcement, similarity to the learner, social power, and status envy are factors that seem to make potential role models attractive [13]. Virtual reality may be seen as an ideal medium to create all kinds of successful and powerful role models, as unbounded from moral obligations, conscience and consequences as they might be. The entertainment industry already feeds us with a never ending stream of strong, attractive and successful "heroes" operating outside the law. These are the role models that many young and vulnerable children grow up with, often in broken families without any sound alternative role models. What kind of virtual reality role models will the children grow up with tomorrow?

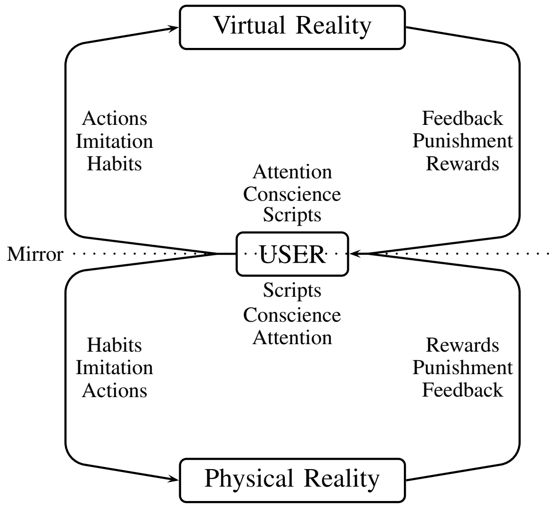


Fig. 2. Parallel learning cycles in a world of mixed realities. Note that the double nature of the on-going moral looping might confuse and affect the morality of the user.

III. Ethical implications

The ethical analysis that follows is built around the applications virtual reality gives or can give rise to. Central for the discussion is the previously mentioned powerful and active learning environment virtual reality constitutes, which is also illustrated in Figure 2. The attentive user learns through repetition and reinforcement, imitation and feedback. Since the whole situation mimics or mirrors the real world, there is always a risk the user gets confused in keeping track of what is real and hyperreal. The parallel learning cycles cause a moral looping inside the users mind giving rise to a set of two different normative ethics speaking differently of what is right and wrong, one for the real, and one for the virtual. Under certain conditions, the distinction between the two might get blurred or it may even disappear, maybe just for a fraction of a second.

This is when things can get really dangerous. Repetitive rewarding actions in the virtual space teach the user social scripts, which automatically can get activated under similar situations in real life much as a stimuli-response instinct. This is a well-known phenomenon. For example, in the army, soldiers with impressionable minds (i.e. adolescents or even children) are trained to kill in combat instinctively by bypassing or short-circuit the reflective mind. This training is known as operant conditioning, and it has proven to be a highly efficient method to improve soldiers skill and will to kill [13]. The tools used in the conditioning program sound familiar. The attentive subjects go through repetitive cycles of stimuli, response and rewards, as they learn the scripts, not seldom in virtual game-like combat fields. Why? Simply because the military forces have realized the educational potential of this technique.

Hyperviolent virtual reality like combat, or hunt and kill, video games are used in millions of homes today and they offer much of this same potential to our young civilians. What are these games doing to them? Are they teaching them to think morally different from previous generations? Are they conditioned? For example, in real life, adults rarely demonstrate how to violently attack other humans in hurtful ways, and then encouraging the children to have a go, but virtual characters can teach attack strategies and allow and encourage the users to refine their capabilities repeatedly, all in a very realistic setting.

Obviously people can get hurt from virtual experiences. It is even true that sometimes people — with or without virtual reality — lose grip on reality. The important question is not whether virtual reality can influence and condition people, but rather how this powerful medium should be handled.

Sometimes voices are heard arguing that the users of virtual reality are capable of telling the difference between actions in the real and the virtual world. For example, simulations filled with rough and brute virtual violence are considered as pure entertainment, and the users are perfectly aware of this innocent setting, and thus they are left pretty much unaffected. It is just a game, right?

Clearly, this argumentation obfuscates research findings. Just because a user can tell the difference between the real and the virtual, it does not follow that he is left unaffected and unhurt. It sounds ridiculous to say that a person spending a lot of time playing, for example, a highly violent virtual reality game would be exactly the same person afterwards. What we take in affects us, in one way or the other, regardless of whether it is gritty dialogue, or acts of violence, and whether it is virtual or real. For several reasons, this should not come as a surprise, as Bivins and Newton also states [5]:

> Advances in neuroscience even suggest that unconscious processing of perceptual stimuli may guide our behaviors. Thus, the journey from the phenomenal to the noumenal is a round-trip ticket. We take with us the strength of our moral convictions to serve us on our journey and bring back the spiritual expansion that always comes of travel.
>
> Virtual experiences become known to us through physical perceptions of sight, sound, touch, taste, and smell and, therefore, cannot help but affect our so-called real selves. Hence, virtual exploration of an activity that might be deplorable in real life, such as rape or murder, cannot help but shift—for better or worse, minutely or grandly—the integrity of the actor's being.

Increasingly realistic virtual experiences will of course make the users less aware of the difference between the virtual and the real. To illustrate and extend the discussion of experiencing mixed realities, with corresponding mixed moralities, and the potential confusion arising from it, a virtual reality sub-field called augmented reality can be considered. In an augmented reality system, the user sees a real scene augmented with additional computer generated information, with the purpose of enhancing the users abilities and perception. In this case, the ultimate goal is seamless integration; that is, to create such a convincing experience that the user cannot distinguish the real world from the virtual augmentation of it[2]. Since the user is not completely immersed by the system, but instead, through the view of the real scene maintains a sense of presence in the real, the superimposed augmentation is perceived and experienced in a real context. This part real, part hyperreal, paradigm, in particular when refined towards it goal, will literally let the user perform action in mixed realities, and if a participant cannot tell the difference between the real and its augmentation, there simply cannot be any perfectly aware user of which is what.

Furthermore, how easy we can be affected of what we see, even as laid back TV watchers, is apparent form the massive load of TV commercials that literally floods our living rooms. Why would the driving market forces repeatedly pay huge amounts of money for a few seconds of TV time, if it was not a highly successful market strategy? Their sole purpose of buying time in the air is to maximize their own profit, and we are the impressionable consumers. As we have argued, virtual reality offer a much stronger potential for influencing people.

A. *Holo-crimes*

In the Star Trek series, a virtual reality simulation system called the holodeck is available for the crew members' convenience. The holodeck is a reprogrammable device that provides highly realistic computer-generated artificial worlds to the participants. It can be seen as the ultimate virtual reality theater in which crew members can visit and experience another time and place that will look, feel, sound, and smell like the real thing. Furthermore, the accuracy of the simulation makes it impossible for the participants to tell holo-humans from real people, and every aspect and detail of any actions carried out by the participants is simulated into perfection.

Although the holodeck exists only in science fiction, it can be considered as the ultimate goal of virtual reality research, which takes small steps towards this goal continuously. It can also be considered as the ultimate video game machine. Every year the video game manufacturers try to improve the realism as much as possible in their game devices. For the sake of the argument, it is therefore meaningful to consider highly realistic virtual reality experiences, as had they been experienced on the holodeck, and their consequences. For example, what is the ethics of holo-assault, holo-murder, holo-genocide, holo-rape, and holo-pedophilia?

Based on a gut feeling, most of us would consider such virtual actions wrong, even if no human is hurt directly. Regardless of any future consequences, the acts seem to be morally objectionable and disgusting in themselves. "Some acts are simply not acceptable even in private" [22]. This implies that people have a moral duty not to hurt themselves or find pleasure and fascination in images of depraved and criminal acts. In particular, children must be protected from

[2]In the field of photorealistic augmented reality, virtual objects are inserted in real scenes by a reconstruction of the lighting conditions, followed by a light simulation for new augmented images.

things that endanger their mental health and development towards safe, sound and responsible citizens.

In the old classical Aristotelian theory of ethics, the importance of forming virtuous habits, in ourselves, from our early childhood, is recognized. Through education, reason, and habits, we must cultivate our character. By taking part in simulated acts of violence, we do harm to ourselves by cultivating bad habits and the wrong sort of character [17].

Thus, a severe objection against these types of single-person simulation experiences is the bad influence it will have on the person. How is the person affected mentally and morally? What if a person goes on to do it for real? As previously mentioned, the active learning environment that virtual reality constitutes makes this fear and objection strong. Research on the effects of media violence has established several harmful short term and long term effects. Furthermore, several horrifying schoolboy shootings with absolutely devastating consequences in both the US and Europe have a suspected coupling to obsessive playing of highly violent video games.

Virtual sex and pornography are also related to virtual violence. Many people are alarmed at the possibility that virtual pornography promotes and encourage pornography in real life, in particular in relation to child abuse. Can holo-rape and holo-pedofilia, as a form of re-creation, entertainment, or desire, really be morally defended? Horsfield explains that

> Those promoting representational pornography as safe virtual experience argue that it is quite different from actual abusive behaviors toward women. Feminist critics legitimately point out, however, that representational pornography is neither fiction, nor virtual: Actual exploitation and violence of women occur in the production of the representations. Likewise, the fictional narratives constructed in pornography so closely parallel actual behaviors experienced by many women that one cannot say that pornography is fictional. If aspects of the content of a constructed virtual environment mirror the power relationships of the actual world, for whom is it 'virtual?' [14]

Surprisingly, it seems as the US Supreme Court in April 2002 has ruled that entirely computer-generated virtual child pornography is protected free speech. In any case, regardless of whether such virtual child pornography does any harm to real children or not, Levy argues that it necessarily eroticizes inequality and contributes to the subordination of women [16].

For the reasons discussed above, holo-crimes seems morally wrong. The often-heard counter argument is based on the old Aristotelian notion of catharsis. According to the catharsis hypothesis or "venting", the violence in society would decrease if individuals could behave in a symbolically aggressive way as a replacement for more harmful forms of aggression. This venting can take the form of active actions, as is the case when, for example, a violent video game is played. Alternatively, the symbolic behavior can take the form of observing aggressive actions, for example, watching a violent movie or aggressive contact sports. One can also imagine people arguing that sexual abuse might actually decrease in society through a cathartic effect, if mentally disturbed sex abusers would be allowed to act out their depraved and perverted wishes and fantasies through virtual experiences.

Clearly, arguments like these are very dangerous. Despite the widespread belief in the catharsis hypothesis, it is contradicted by the results of several recent research studies [8]. In fact, the opposite effect is produced; that is, the "venting" stimulates further aggression and actions. In the testimony given by psychology researcher Anderson, in the senate hearing 2000 on the impact of interactive violence on children, he lists the catharsis hypothesis as one of the myths around us concerning media violence, also stating that behaving aggressively or watching aggressive behavior increases subsequent aggression [1].

Another raised concern is that many people might tend to prefer the virtual over the real, potentially leading to social decay, where people start to ignore their personal and civic duties. In the extreme case, virtual reality becomes the "ultimate opiate", where the participant spends all their awake time creating successful and powerful alter-egos and accomplish things they would never be capable of in the real world, which they might also regard as a too unfair and hostile place for them to live in [22], [14].

B. Human rights

There are several important human rights and liberties that we need to defend in our society. For example, the importance of free speech legislations like the First Amendment in the US must be recognized:

> Congress shall make no law respecting an establishment of religion, or prohibiting the free exercise thereof; or abridging the freedom of speech, or of the press; or the right of the people peaceably to assemble, and to petition the Government for a redress of grievances.

Proponents for controversial video games often argue that, regulating or banning the production or the sale of "bad art" has nothing to do with a free society. It should be noted, however, that the right to free speech is often used for different purposes than for what it was originally intended. In particular, the market actors seem to have rephrased the right to freedom of speech into their right to maximize profit on a free, unregulated market.

Products endangering our own health, or the health of others, must be dealt with according to the "principle of precautionary action". The possible negative consequences of virtual reality include addiction, increased aggressiveness, and other medical and psychosocial conditions. Therefore, the libertarian viewpoint, stating that the decision to engage in holo-crimes, or not, is entirely a private consideration, is troublesome. Sometimes the individual freedom must be limited, for the well-being of others.

Furthermore, according to the UN child convention, children have the right to healthy development, both physically and mentally. This right must be protected first and foremost. The sale of mentally harmful material to children can therefore never be considered to be constitutionally protected. This makes regulation of the video game, and the emerging virtual reality entertainment, market, an appropriate alternative.

Finally, not all sorts of communication is protected free speech. Discrimination of women, or minorities, based on, for example, race or religion, is banned in all civilized societies. There are legislations trying to prevent such discrimination. In the same spirit, discriminating virtual reality entertainment must be dealt with.

C. Designers' responsibility

The designers of a virtual reality determine the set of available user options and actions. Thus, the consequences of actions are in the hands of these creators or virtual parents. They decide whether morally reprehensible acts are possible, and whether such behavior is rewarded or punished. In this way, the users are guided in their choices of actions. If only a very limited set of options is available, the user effectively becomes a marionette in the hands of the designer.

Since the designers of a virtual reality application are in a position where they can signal social approval or disapproval for the actions made possible, they carry a moral burden on their shoulders to uphold high moral standards. This includes the way they choose to represent behavioral options and consequences of actions [22], [6]. If legitimacy is given to reprehensible acts, like maiming, murders, rape, and discrimination in virtual socials context, through various kinds of encouragement and rewards, the designers act unethically. In fact, they betray the users by effectively teaching them that ethics is schizophrenic or superfluous.

Note that it has not been argued here that all applications need to be free from violence or criminal themes. To convincingly resemble real human life, the possible courses of action in a virtual reality must represent a broad spectrum of what is morally good and bad. The user must be allowed to act as free agents, but also forced to take responsibility for their actions, which includes punishment of criminal acts.

For example, applications in which the user is allowed to see things from the victims point of view, or where the virtual criminals get proper punishment, cannot be seen as controversial as, for example, the mandatory excessive use of violence and murder exercised by glorified heroes (which most often are controlled by the user himself) operating outside the law in today's prevalent hyperviolent computer games.

To help virtual reality designers to take social responsibility, a code of ethics with moral imperatives might be needed, developed in a similar spirit as, for example, the ACM code of ethics for computer professionals. Hopefully, by agreeing upon imperatives aiming at preventing harmful effects of the medium, and making sure the users needs are clearly assessed during system development, ethical principles and reflections becomes a more natural part in the development process.

From a social perspective, ethical issues also arise concerning the construction of and access to the medium, in particular since media and technology are central tools for people in power [14]. Whose world views are reinforced? Whose political and economic interests are covered? Who has access to the needed technology? What is presented, what is excluded, and why?

IV. Conclusions

As we have discussed, highly realistic virtual reality simulation of criminal acts have the potential to severely affect the participants, as well as others indirectly. Applications encouraging virtual violence or pornography, glorifying or giving misleading or unrealistic views of such phenomena, are therefore morally objectionable and unethical. The hyper-violent realistically looking computer games that already hit the shelves in millions of copies are a clear indication of what type of depraved applications we can expect from future virtual reality platforms.

It is therefore imperative that virtual reality developers, educators, cultural workers, and parents choose a precautionary and conservative line with respect to the virtual realities introduced in our society. The ethical problems raised by the applications must be addressed both pedagogically and politically. For example, educational programs in our schools are needed that discuss the ethical implications of new advanced media. Legislators must also ensure that children and adolescent can enjoy a proper protection from the possible harmful effects. For example, it might be a good start to extend the age-based censorship in use in many societies for movies to computer and video games, virtual reality entertainment included.

With or without regulation, it should be clear that a high responsibility rest on the shoulders of the virtual reality developers and manufacturers. In all their work, they ought to reflect upon its meaning, purpose, and consequences. High ethical standards must also be established to support and guide them. One possible way would be to enact an obligatory ethics education law for the people employed in the industry, giving them a license to work in the field.

Finally, more research is needed to clarify the problems involved and to make well-informed decisions possible. Where shall we draw the line between the acceptable and unacceptable, the thinkable and the unthinkable? How do children and adolescent respond to various types of applications? What type of virtual reality experiences can have a good and prosocial influence on the participants? Since virtual reality is still in its infancy, our ethical reflections must also be enhanced and updated as the medium is further developed. Most certainly, as the technology is improved, it will become a prevalent tool in many more areas. Novel and revolutionary applications, utilizing a fuller range of the medium, will enter the scene.

Although the discussion herein have been centered around the potential degree of similarity of the real and the hyperreal, and the educational and ethical consequences of this, it is of course possible to use virtual reality to explore areas beyond the objective world. The results and ethical concerns this will lead to are interesting endeavors worthy future discussions. Hopefully, applications can be created leading to new insights and understanding for the benefit of mankind, perhaps even in the area of morality itself?

References

[1] C. A. Anderson. Violent video game increase aggression and violence. U.S. Senate Commerce Committee hearing on "The Impact of Interactive Violence on Children", March 21, 2000.

[2] C. A. Anderson and B. J. Bushman. Effects of violent video games on aggressive behavior, aggressive cognition, aggressive affect, physiological arousal, and prosocial behavior: A meta-analytic review of the scientific literature. *Psychological Science*, 12(5):353–359, September 2001.

[3] A. Bandura. *Social Foundation of Thoughts and Action: A social Cognitive Theory*. Englewood Cliffs, NJ: Prentice Hall, 1986.

[4] J. Baudrillard. Simulacra and simulations. In M. Poster, editor, *Jean Baudrillard, Selected Writings*, pages 166–184. Stanford University Press, 1988.

[5] T. H. Bivins and J. H. Newton. The real, the virtual, and the moral: Ethics at the intersection of consciousness. *Journal of Mass Media Ethics*, 18(3&4):213–229, 2003.

[6] P. Brey. The ethics of representation and action in virtual reality. *Ethics and Information Technology*, 1:5–14, 1999.

[7] B. J. Bushman and C. A. Anderson. Violent video games and hostile expectations: A test of the general aggression model. *Personality and Social Psychology Bulletin*, 28(12):1679–1686, December 2002.

[8] B. J. Bushman, R. F. Baumeister, and A. D. Stack. Catharsis, aggression, and persuasive influence: Self-fulfilling or self-defeating prophecies? *Journal of Personality and Social Psychology*, 76(3):367–376, 1999.

[9] S. L. Calvert and S.-L. Tan. Impact of virtual reality on young adults' physiological arousal and aggressive thougts: Interaction versus observation. *Journal of Applied Development Phsychology*, 15:125–139, 1994.

[10] K. E. Dill and J. C. Dill. Video game violence: A review of the empirical literature. *Aggression and Violent Behavior*, 3(4):407–428, 1998.

[11] J. B. Funk, D. D. Buchman, J. Jenks, and H. Bechtoldt. Playing violent video games, desensitization, and moral evaluation in children. *Applied Developmental Psychology*, 24:413–436, 2003.

[12] W. Gibson. *Neuromancer*. Ace Books, 1984.

[13] D. Grossman. *On Killing: The Psychological Cost of Learning to Kill in War and Society*. Back Bay Books, 1996.

[14] P. Horsfield. Continuities and discontinuities in ethical reflections on digital virtual reality. *Journal of Mass Media Ethics*, 18(3&4):155–172, 2003.

[15] J. S. Kollin. A retinal display for virtual-environment applications. In *Proceedings of Society for Information Display*, page 827, 1993.

[16] N. Levy. Virtual child pornography: The eroticization of inequality. *Ethics and Information Technology*, 4:319–323, 2002.

[17] M. McCormick. Is it wrong to play violent video games? *Ethics and Information Technology*, 3:277–287, 2001.

[18] R. Schroeder. Playspace invaders: Huizinga, Baudrillard and video game violence. *Journal of Popular Culture*, 30(3):143–153, 1996.

[19] D. Stanovsky. Virtual reality. In L. Floridi, editor, *Philosophy of Computing and Information*, pages 167–177. Blackwell Publishing, 2004.

[20] I. Sutherland. The ultimate display. In *Proceedings of IFIPS Congress*, pages 506–508, 1965.

[21] K. Warwick. Cyborg morals, cyborg values, cyborg ethics. *Ethics and Information Technology*, 5:131–137, 2003.

[22] B. R. Withby. The virtual sky is not the limit — the ethical implications of virtual reality. *Intelligent Tutoring Media*, 3(2), 1993.

When simulations become reality

Magnus Johansson
Department of Information Technology
Uppsala University
Sweden

***Abstract*— When creating models of some phenomenon in the real world we often strive for perfection. We want the model to closely resemble the real world so that we can perform virtual experiments on it without risking our physical health or the environment. In this paper we discuss what kind of impact on our understanding of reality such perfection would have and how we could distinguish perfect simulations from reality, if at all.**

I. Introduction

In today's society computer simulations have replaced many dangerous or expensive tasks and experiments. In many cases people strive to make the model as close as possible to the reality it mimics. For example, a flight simulator tries to mimic a real flight experience so that the pilots get good training. A perfect flight simulator would end the need for pilots to practice in real planes, saving lots of money.

You may also want to create a model or a simulation for a different reason. Instead of trying to get as close as possible to the reality you may want to simplify it. You abstract away from details in order to be able to focus on the more interesting part of the phenomenon you are studying. For the purposes of this paper, this kind of model is not as interesting as the kind which tries to be as real as possible so we do not discuss it here.

In this paper we discuss if we can create a perfect simulation of our reality, a kind of simulation of everything, and some of the implications on our understanding of reality such a simulation would have. In section II we give a few definitions that we will use subsequently. Section III describes some examples of perfect simulations in science fiction. In section IV we define a test that can be used to determine whether or not a simulation is perfect. Section V raises some questions about death in simulations, and section VI argues that if we can create a perfect simulation, we will be able to recursively create new simulations. Finally, section VII discusses some related work.

II. Reality, models, and simulations

A. Reality

In order to talk about models of the reality we need to have some kind of idea of what reality is. According to the dictionary [1], reality is ”something that is neither derivative nor dependent but exists necessarily”. Another dictionary [2] has a similar definition: ”The totality of what is, as opposed to what merely seems to be.”. The problem with these definitions is that they define reality either in terms of existence (”exists necessarily”, ”what is”) or in terms of non-existence (”seems to be”). These definitions therefore depend on a definition of existence. According to [2], existence is ”Instantiation in reality, or actual being.”. Hence, it is a circular definition.

If you look to history for ideas of what reality is you will get many different answers. The following is from Heim [3]:

> Plato holds out ideal forms as the ”really real” while he denigrates the raw physical forces studied by his Greek predecessors. Aristotle soon demotes Plato's ideas to a secondary reality, to the flimsy shapes we abstract from the really real – which, for Aristotle, are the individual substances we touch and feel around us. In the medieval period, real things are those that shimmer with symbolic significance. The biblical-religious symbols add superreal messages to realities, giving them permanence and meaning, while the merely material aspects of things are less real, merely terrestrial, defective rubbish. In the Renaissance, things counted as real that could be counted and observed repeatedly by the senses. The human mind infers a solid material substrate underlying sense data but the substrate proves less real because it is less quantifiable and observable. Finally, the modern period attributed reality to atomic matter that has internal dynamics or energy, but soon the reality question was doomed by the analytical drive of the sciences toward complexity and by the plurality of artistic styles.

Just as Heim concludes, for two thousand years Western culture has puzzled over the meaning of reality.

However, everybody has an intuition of what reality is, and we will simply use that intuition when talking about reality. Reality is what you perceive as real.

B. Models versus simulations

So far we have used the terms *model* and *simulation* without specifying what we actually mean by them. It is hard to find one definition of model and one definition of simulation. Different authors define the terms in different ways, if at all. Often the terms are used interchangeably. However, in a specific domain, one of the terms usually dominates (e.g. flight *simulations* and economic *model*). For the purposes of this paper we will use the term *model* and *simulation* in the following fashion:

Definition 1 (Model): A static representation of a phenomenon.

Definition 2 (Simulation): A running model (a model ”put to live”).

A model is something static. You can look at it and see what properties it has and how it is constructed. A simulation, on the other hand, is something dynamic, a *running model*. For example, the traffic planning model is a model consisting of mathematical formulas describing the traffic flow in the intersection. When you start using the model by instantiating it with some data and calculating subsequent states of traffic according to the formulas, the model becomes a simulation.

C. Immersive simulations

We will distinguish between two kinds of simulations. We will call them immersive and non-immersive, respectively. An immersive simulation is one where you are part of the simulation, like the flight simulator or when you drive the model of the vintage car. The flight simulator tries to trick you into believeing you are flying an actual plane, and the model of the vintage car is trying to make you believe you are driving the original. Non-immersive simulations are simulations where you are not part of the simulation. You may still control the simulation, but you are not a part of it. The car crash test example is a non-immersive simulation. In this paper we will focus on immersive simulations.

D. Perfect simulations

As mentioned in the introduction, when we create a model or a simulation of some real world phenomenon we strive for perfection. We want the simulation to behave as close as possible to the real world. Things that can prevent us from reaching perfection are e.g. economy and technology, but still we try to reach it. It is, after all, the point of the simulation to mimic the real world as close as possible. But how close is perfect? It depends on what you are simulating. For the car crashing example it would be that all possible car crashes in all possible ways with all possible car combinations in all possible environments can be simulated. Can it be achieved? Perhaps. For the vintage car example it would be that the car is indistinguishable from the original. Can it be achieved? Likely. There are still people questioning the authenticity of the Mona Lisa in the Louvre (see e.g. [4]). A perfect flight simulator would make the pilot believe he is piloting an actual airplane. Perfection is model dependent. The criteria for perfection for one simulation need not be the criteria for another simulation. Perfection lies in the eye of the beholder. What is perfect for one observer may not be perfect for another. We will use the following definition of a perfect simulation:

Definition 3 (Perfect simulation): A simulation is perfect when you can not tell it apart from the phenomenon it simulates.

E. The universal simulation

We have all these simulations of different aspects of the real world. The crash simulator simulates how cars crash, the weather simulator simulates how the weather behaves, and the flight simulator models an airplane etcetera. Can we create one immersive simulation that encompasses all other simulations? That is, one simulation that simulates car crashes, traffic, weather, and all other real world things we can imagine? Let us assume that we have such a simulation and let us call it *the universal simulation*.

III. Immersive universal simulations in fiction

There are examples of immersive universal simulations in science fiction. The two most well known are the holodeck from the TV-series Star Trek, and the matrix from the movie with the same name. We will take a brief look at them.

A. The holodeck

The holodeck is a room with special properties. You immerse yourself into the simulation simply by entering the room. There are actually two different technologies described that explains how the holodeck works, one from the series Star Trek: The Next Generation (TNG), and one from Star Trek: Voyager (VOY).

In TNG the holodeck is a room where the walls, the floor, and the ceiling can generate holographic images that appear to stretch out in the distance. Holograms can also be projected into space to give impression of objects closer to the user. When there is to be a physical interaction between the user and a hologram, the hologram is augmented with force beams. Things that need a more intimate interaction with the user are replicated[1], e.g. food and water. If you walk around in the holodeck you will sooner or later bump in to a wall. This problem is solved by equipping the holodeck with a tread mill of force fields. This way the holodeck can shift the user around to prevent him from hitting a wall. Using gravity modifications the user will not notice the movement.

In VOY the holodeck is still a room with holographic walls, but when interaction is about to take place between the user and some holographic object something called magnetic bubbles are used instead. These bubbles are molecule sized and can be controlled in three dimensions by the computer controlling the holodeck, giving the sensation of a real object.

More information about the holodeck can be found in [5].

B. The matrix

The matrix takes a different approach to immersing the user into the simulation. Instead of a real world room to enter, a neural interface is used. You get into the simulation by connecting the neural link embedded in your skull to the computer running the simulation. The computer then feeds the brain with information from the simulation, giving you the impression of real life. While in the simulation your real life body remains immobile. Signals from the brain to your muscles are intercepted by the neural interface and translated to simulated movement in the simulation.

An interesting difference between the matrix and the holodeck is that in the holodeck you can eat real food, since your whole body is inside the simulation and food is being replicated on demand within the simulation. In the matrix, on the other hand, your brain is only given the impression of eating. Your real life body still needs to receive nutritions if you are to stay in the matrix for any extended period of time.

[1] A replicator is essentially a xerox machine for real world things. It copies items down to the molecular level.

IV. Test for reality

Is the holodeck or the matrix a perfect immersive universal simulation and what do we mean by perfect in this case?

In his 1950 paper ”Computing Machinery and Intelligence” [6], Alan Turing put forward the idea of an imitation game to decide whether a computer was intelligent or not. The idea is that you put an interrogator in front of a device and let him/her ask questions to it by entering textual messages. If the interrogator can not say if the device is controlled by a human or a computer, then, Turing argues, it would be unreasonable to not call the computer intelligent. Others have argued that this test is not sufficient because of the limitations of the textual interface. For example Harnad [7] argues that the computer needs to be augmented with a body of some kind in order to determine whether or not it is intelligent. The purely textual interface is not enough because a computer might pass this test without really being intelligent. The new test requires the computer to fool the tester, not only linguistically, but also with body language and all interactions with the environment that a human can do. Harnad calls this new test ’the Total Turing Test’. Schweizer [8] takes this test even further. He argues that the Total Turing Test is not sufficient because the programmer of the computer (or robot) can preprogram all information about the environment in which the computer is to be tested. A truly intelligent being would not have to have that information. It would be able to aquire it on its own. Schweizer argues that the robot needs to be observed over an extended period of time to see how it behaves in an arbitrary environment. If its behaviour cannot be distinguished from that of a human, the robot is intelligent.

Let us create a similar test to decide whether we are in an universal immersive simulation or in the real world. Let us say you go to bed at night, and during your sleep someone has the ability to immerse you into a universal simulation and/or move you to a different place. You do not know whether or not this someone actually will immerse or move you. When you wake up you can do whatever you want and use whatever instruments you have or can create to try to decide if you are in an immersive universal simulation or in the real world. If you can not do so and you actually are in a simulation, the simulation is perfect. Then what grounds do you have to call the perfect universal immersive simulation unreal? As far as you can tell it is the reality. It fits your intuitive idea of what reality is.

In response to Nick Bostrom [9], John D. Barrow argued [10]

> ...if we live in a simulated reality we should expect occasional sudden glitches, small drifts in the supposed constants and laws of Nature over time, and a dawning realisation that the flaws of Nature are as important as the laws of Nature for our understanding of true reality.

According to this reasoning, if we live in a simulation we should be able to find glitches that gives away the fact that we live in a simulation. We should start looking at phenomena we cannot explain by the laws of Nature.

But how can you tell that the glitches you observe are not part of the laws of Nature that you just have yet to understand? Barrows argument is based on an assumtion that you can tell glitches and drifts apart from laws of Nature. This might be true if we compare a simulation we created to our reality. Then the laws of Nature to compare against would be the laws of Nature as we know them in our reality. The glitches in the simulation would then be the places where the laws of Nature in the simulation differ from the laws of Nature in our reality. Lacking a reference reality we cannot tell whether or not any ”glitches” we see actually are glitches, or just a part of the laws of Nature that we have yet to understand.

A. How can we test anyway?

So what would you do to find evidence of living in a simulation? You could learn from software programming, by looking at the extreme cases. Often glitches (or bugs) are found in the extreme, unanticipated cases. You could perhaps start looking at the very small and try to find glitches. Maybe quantum physics as we know it is evidence of a glitch. Or you could start looking at the very big. Maybe the Big Bang is evidence of a glitch.

Attempts like these will most likely fail because we have no means of telling whether the results are glitches or something that is supposed to be part of our reality. Are there any other ways to proceed? Heim [3] gives an indication of where we should look. He describes three *anchors* that anchors us to the real world: mortality/natality, carryover between past and future, and care. We are born at a definite time and we die at a definite time. He argues that ”these limits impose existential parameters on reality, providing us with a sense of rootedness in the earth”. As for the carryover feature he argues that it ”distinguishes reality from any passing entertainment of momentary hallucination”. Finally, by care he means that we are fragile and therefore need to care. He claims that ”these three features mark human existence and stamp experience with degrees of reality”.

He does however look at them from a different perspective. He sees these features as something to avoid to some extent in virtual reality:

> Should synthetic worlds, then, contain no death, no pain, no fretful concerns? To banish finite constraints might disqualify virtuality from having any degree of reality whatsoever. Yet to incorporate constraints fully, as some fiction does, is to produce an empty mirror over and above the real world, a mere reflection of the world in which we are anchored. [...] Actual cyberspace should do more; it should evoke the imagination, not repeat the world. Virtual reality could be a place for reflection, but the reflection should make philosophy, not redundancy.

In a perfect simulation all these features would exist, and the simulation would be ”an empty mirror”. But that empty mirror of redundancy *would* make philosophy. It would give us insights into the essence of our own reality.

Carryover and care would be relatively easy to implement in a simulator, but how would you deal with birth and death? Maybe the ultimate test is death itself.

V. Universal immersive simulations and death

Depending on how you immerse yourself into the simulation different things would happen if you were to die while in the simulation. If you immerse yourself Star Trek- style, death would be no different from the real world. You still only have one body. If you instead immerse yourself matrix-style with a neural interface, you would suddenly have two bodies. One in the real world and one in the simulation. If your simulated body dies in the simulation, what would happen to the real body? Either that dies as well, which is what happens in the matrix, or it will continue to live, perhaps ”waking up” from the simulation. A more interesting scenario would be if your real body dies (perhaps from starvation) while you are in the simulation. What would then happen to your simulated body? What if you share the simulation with others? How would they experience your real body dying in the simulation?

I will not try to answer these questions, but it is interesting to note that many religious beliefs can be explained in terms of simulations like this.

VI. The recursive nature of perfect universal simulations and reality

Models are built from parts of the real world. We create refined materials like plastics, silicone, and metal, and from those we build computers, and with computers we create models and simulations. When trying to build the perfect immersive universal simulation we will either succeed or fail. If we succeed we will end up with a simulation, s_1, in which we would also have the ability to create a new perfect immersive universal simulation, s_2. If we are not able to do so, s_1 would not be perfect because we could test it by trying to build s_2. If we fail we know we are in the simulation s_1, contradicting our assumption that we succeeded building a perfect s_1. If we have a perfect s_1, that is, we have a model that we can not distinguish from reality, can we know if our reality is *the* reality or just another perfect simulation? Is there a base case for this recursion, a $reality_0$?

If we fail building s_1, however, we will still not know anything more about our reality since we can not exclude the possibility that we failed because we are not yet good enough, and not because it is impossible. Maybe the best way to learn more about our reality is to try to create a perfect universal model of it, and succeed. Or fail and understand why we failed. Heim [3] has a similar idea:

> Rather than control or escape or entertain or communicate, the ultimate promise of VR may be to transform, to redeem our awareness of reality – something that the highest art has attempted to do and something hinted at in the very label *virtual reality*...

VII. Related work

Nick Bostrom argues in [9] that there is a real chance that not only are we living in a simulation, but we are also simulations. He argues that at least one of the following propositions is true: (1) The fraction of human-level civilizations that reach a posthuman stage is very close to zero; (2) The fraction of posthuman civilizations that are interested in running ancestor-simulations is very close to zero; (3) The fraction of all people with our kind of experiences that are living in a simulation is very close to one. Using the terms presented in this paper this can be explained as follows. Either

1) we will never be able to create a perfect universal simulation before going extinct, or
2) we will not use the technology, if we get it, to simulate our ancestors, or
3) if we can simulate our ancestors, then it is very likely that we are being simulations ourselves.

VIII. Conclusion

Testing whether our reality is a simulation or not will prove very difficult without some kind of reference to test against. Simulations we create could be easily tested since we have our own reality as a reference, but testing our reality will be much more difficult. Maybe the successful creation of a perfect immersive simulation could provide us with such a reference and give an indication of whether or not we live in a simulation.

References

[1] (2004) Merriam-webster online. [Online]. Available: http://www.m-w.com/

[2] G. Kemerling. (2002) A dictionary of philosophical terms and names. [Online]. Available: http://www.philosophypages.com/dy/index.htm

[3] M. Heim, *The Metaphysics of Virtual Reality*. Oxford University Press, 1993.

[4] (2004) The proliferation of forgery. [Online]. Available: http://www.encyclopedia.com/html/section/forg-art_TheProliferationofFor%gery.asp

[5] J. Bell. (2003) Holodeck and computers faq. [Online]. Available: http://www.calormen.com/Star_Trek/FAQs/holodeck-faq.htm

[6] A. Turing, “Computing machinery and intelligence,” *Mind*, vol. 59(236), pp. 433–460, 1950.

[7] S. Harnad, “Other bodies, other minds: A machine incarnation of an old philosophical problem,” *Minds and Machines*, vol. 1, pp. 43–54, 1991.

[8] P. Schweizer, “The truly total turing test,” *Minds and Machines*, vol. 8, pp. 263–272, 1998.

[9] N. Bostrom, “Are you living in a computer simulation?” *Philosophical Quarterly*, vol. 57(211), pp. 243–255, 2003. [Online]. Available: http://www.simulation-argument.com

[10] J. D. Barrow. (2003) Living in a simulated universe. [Online]. Available: http://www.simulation-argument.com

Early Stages of Vision Might Explain Data to Information Transformation

Baran Çürüklü
Department of Computer Science and Engineering, Mälardalen University, Västerås S-721 23, Sweden
Baran.curuklu@mdh.se

Abstract

In this paper we argue that the process of attaching meaning to data, in order to produce information, occurs at first within the early stages of the sensory areas. This hypothesis is based on an analysis of the signal flow within the early stages of vision. It is evident that the simplest form of visual information, which is an edge, is conceptualized at first within the primary visual cortex. Support for this hypothesis comes also from computational vision theories, since the simplest features that can be extracted from an image are closely related to what is represented within the primary visual cortex.

1. Introduction

It has been suggested that information is the result of attaching meaning to data [3,6,7]. According to this view data is transformed, through a chain of processes, to generate information. However, this view does not deal with the question of where, i.e. in which sub-structure of the brain, is data transformed into information. Obviously, such a transformation involves neurons, which through interactions with each other represent our memory and enable us to interact with the environment. A justifiable question is if it is possible to locate those neurons or sub-structures of the brain that lead to the data to information transformation.

The primary visual cortex (V1) is the end station of the early stages of vision and interacts heavily with higher visual areas (Fig. 1). Thus, an analysis of the activity that starts in the eyes and continues to the primary visual cortex might reveal where data is transformed to become information. Besides, this analysis is also applicable to other senses as well, since there are structural similarities between various cortical areas, which is primarily manifested by the modular and laminar organization of the neocortex (main part of the brain) [13].

In this paper we address the question of where data is transformed into information. Taking together the evidences from visual neuroscience and computational vision it is plausible to assume that the raw data that enters the eyes are transformed into meaningful information within the primary visual cortex, since the simplest form of information emerges within this structure.

2. Early Stages of Vision

Early stages of vision consist of three structures, i.e. the retina, the lateral geniculate nucleus (LGN), and the primary visual cortex (V1) (Fig. 1). The retina is the first processing step of the incoming signals of the visual field, since the visual field is projected to the two retinae located in each eye. The physiological studies done by Kuffler have shown that the cat retinal ganglion cells react to small spots of light [11] (Fig. 2 *Left*). This study indicates that each retinal ganglion cell responds to stimulation of a small, circular patch of the retina. This patch defines the receptive field of a single retinal ganglion cell. Kuffler has discovered two different types of retinal ganglion cells (ON and OFF) [11]. When the center of an ON retinal ganglion cell is stimulated with a light spot, the cell reacts by generating a burst of spikes as a result of increased activity. When the light stimulus, which covers the center of the ON retinal ganglion cell, is moved to the periphery of the cell's receptive field, the cell is inhibited (decreased activity). The OFF retinal ganglion cells have opposite receptive field properties. These cells prefer dark spots in the center of their receptive fields, which is surrounded by a light region. The receptive fields of the retinal ganglion cells can be mathematically described with a so-called 'Mexican-hat' function.

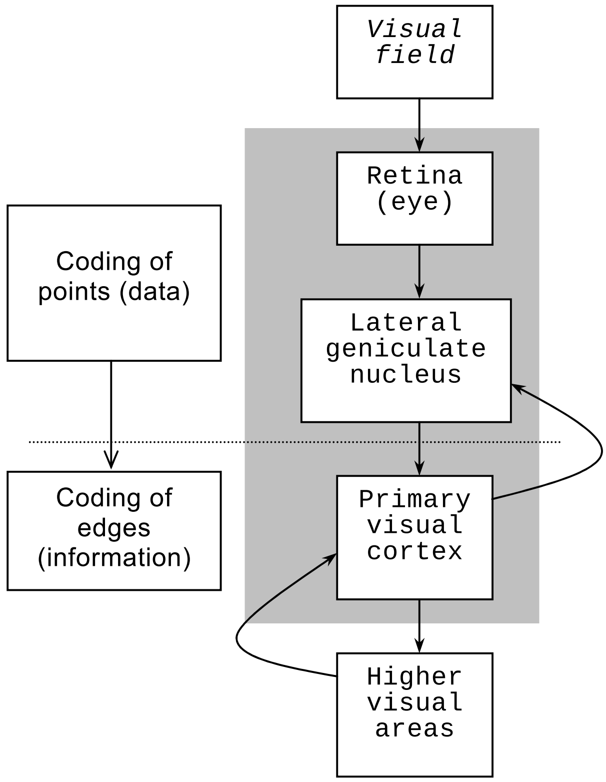


Figure 1. A scheme of the signal flow within the early stages of vision (the gray area). Within the retina and the lateral geniculate nucleus neurons have circular point-like receptive fields. Neurons that populate the primary visual cortex are selective to the orientation of the contrast-edges and lines. The vertical line illustrates the border between the regions that are dedicated to data and information.

Note that the activity of the retinal ganglion cells depends on the contrast, i.e. the difference in the amount of light that falls on their receptive field centers and surrounds. Both types of retinal ganglion cells are equally numbered, and are distributed equally in the retina.

The major target of the retinal ganglion cells is the LGN. Neurons in the LGN send their axons to the primary visual cortex. Note that the retinotopic representation of the visual field, which emerges in the retina, is preserved throughout the retina-LGN-V1 pathway. It has been shown that the fibers that start from the neighboring retinal ganglion cells within the retina converge to neighboring geniculate cells within the LGN [10]. The geniculate cells project in turn to neighboring regions within the V1 [10] (Fig. 2).

Neurons within the LGN are classified as ON and OFF cells similar to the retinal ganglion cells. As a result the LGN is often seen as a relay between the retina and the primary visual cortex. However, the LGN also receives modulatory input from the V1.

Neurons populating the primary visual cortex react selectively to contrast-edge (line) orientations. These cells are named "simple" and "complex" depending on their response properties [9,10]. However, the relay cells within the LGN, which carry the visual signals from the retina to the primary visual cortex, are not orientation selective. It is not known in detail how orientation selectivity of the neurons within the primary visual cortex emerges [1–5,14]. Hubel and Wiesel have proposed that orientation selectivity of the cat simple cells is a consequence of the arrangement of the LGN input [9]. According to this arrangement, the ON-center LGN cells converge to ON-subregions of the simple cells (Fig. 2). The OFF-subregions of the simple cells are constructed in the same way by the OFF-center LGN cells (Fig. 2).

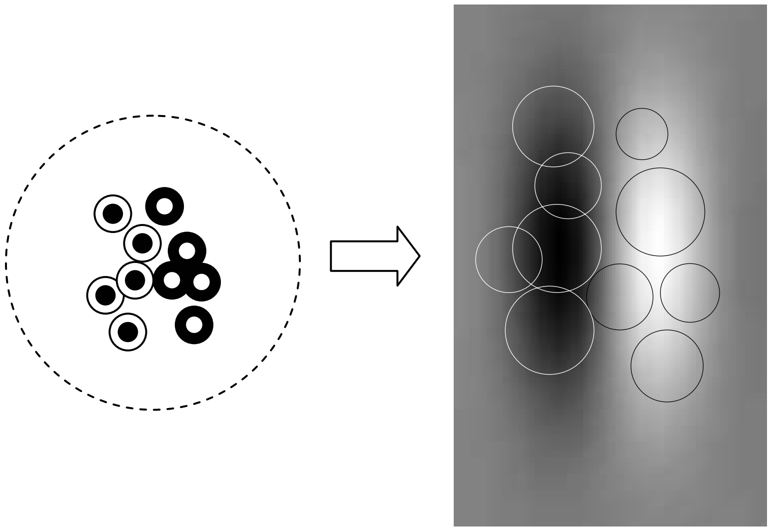


Figure 2. *Left*. An illustration, which shows the receptive fields of ten LGN (or retinal ganglion) cells. Half of them are ON-center and the rest are OFF-center neurons. Other cells that are located within this circular patch are not shown for the purpose of clarity. *Right*. An illustration showing the receptive field profile of a simple cell (located within the V1). Note that the spatial relationships between the LGN (or retinal ganglion) cells are preserved during the projections to the primary visual cortex. The receptive field profile of the simple cell is basically an effect of this preservation

This model is called the 'feedforward' model, since the intracortical connections do not play any prominent functional role in generating the orientation selectivity properties of the neurons within the V1. The signals flow in a pure feedforward fashion, along the retina-LGN-V1 pathway.

The simple cells, which dominate layer 4 of cat, have elongated ON- and OFF-subregions that reflect the LGN input. Hubel and Wiesel have discovered that these cells' responses to complex stimuli can be predicted from their responses to individual spots of lights [9,10]. As a consequence, a simple cell's receptive field can be mapped based on its response to small light spots positioned on different locations on the retina. A light spot, which is positioned at the cells' ON-subregion, excites the cell, whereas a light spot positioned at the cells OFF-subregion inhibits the cell. The responses to 'dark' spots are opposite to light spots.

3. Computational Theories of Vision

The objective of computational vision is to process images, more precisely to detect (or classify) shapes, objects and other primitives. One of the most influential theories within this field is the primal sketch theory by Marr and Hildreth [12]. This theory addresses the detection of lines and contrast edges. Later, it has been shown that the primal sketch theory fits elegantly the Hubel and Wiesel feedforward model of orientation selectivity [9].

According to Marr and Hildreth one can transform a fully analog, grey-scale image to a symbolic representation of image-based features [12]. It was proposed that vision must go "symbolic" right at the beginning. The suggested symbols were bars, contrast edges, terminations and "blobs". Bars (or lines) are short line segments whose terminations lie outside the receptive field. Blobs' terminations lie within the receptive field. Terminations represent "ending of tings".

Another theory of vision is based on the filtering of the images [8]. In this theory it is assumed that an image consists of several sinusoidal gratings with various orientations and spatial frequencies. It is further assumed that neurons that populate the primary visual cortex can be seen as spatial frequency filters. The frequency of such a filter is defined by the width of the ON- and OFF-subregions of the neuron's receptive field. Thus each neuron responds selectively to a narrow band of spatial frequencies. At first sight this way of representing an image seems to be radically different from the primal sketch theory. It is, however, obvious that a contrast-edge has both an orientation and a spatial frequency. It is thus possible to represent a contrast-edge as a sinusoidal grating, which has the width of one cycle. Consequently, it is plausible to assume that the sinusoidal gratings that are detected are as symbolic as the contrast-edges of the primal sketch theory.

4. Discussion

The short overview of the two scientific disciplines shows that the idea of transforming data to produce information is not entirely new, at least not in vision research. Independent of the main objective of these two disciplines, both neuroscientists and computational vision theorists deal with the question of how the systems that they study generate information from incoming data.

It is obvious that, at least within the brain, data is transformed into information continuously (Fig. 1). This transformation process starts at the retina and goes on all the way to the V1, and to the higher visual areas where objects are recognized (Fig. 1). However, feedback connections that emerge within the V1 and target LGN show that the image, which is registered by the eyes, is modified before it reaches the V1. This feedback path is highly interesting since it demonstrates how "more meaningful" data can modify data, which is "raw".

We hypothesize that the feedforward and feedback connections within and between the three structures that are found in the early stages of vision are doing more than passive preprocessing. Note further that when information is defined as "data that has a meaning" it is meant that information must fulfill certain requirements, i.e. information must have a meaning. In terms of vision the simplest form of meaningful information is an edge. An edge can be perceived as a border or a direction. Note that the same cells that analyze orientations of contrast-edges also represent other shapes, such as small points.

Within the early stages of vision the concept of an edge appears first in V1. Recall that neurons in the retina and the LGN analyze simply discrete points within the visual field, whereas neurons within the V1 are tuned to orientations of the edges. This transformation from a set of spatially ordered discrete points to a line is in our view the transformation of data to information.

Similarly, computational vision theories do also stress that fact that vision goes symbolic very early. Data is the pixel values (roughly corresponding to retina and LGN), whereas bars, contrast edges, terminations, blobs and spatial frequencies are the simplest forms of information. Note that all these forms of information emerge within the V1. Consequently, despite differences in approach, both the primal sketch theory and the spatial frequency filter theory assume that data is transformed into information within the V1.

Taking together the evidences from two complementary disciplines within the vision research it is plausible to assume that raw data that enters the eyes is transformed into meaningful information within the primary visual cortex, since the simplest form of information is represented at first within this structure.

5. References

[1] B. Çürüklü and A. Lansner. A model of the summation pools within the layer 4 (area 17). *To be appear at the Annual Computational Neuroscience Meet. (CNS)*, Baltimore, USA, Elsevier. (2004)

[2] B. Çürüklü. Layout and Function of the Intracortical Connections within the Primary Visual Cortex. Licentiate thesis, Mälardalen University Press. (2003)

[3] B. Çürüklü and A. Lansner. An abstract model of a cortical hypercolumn. *In Proc. Int. Conference on Neural Information Processing*, Singapore, pp. 80–85, IEEE Press. (2002)

[4] A. Das. Orientation in visual cortex: a simple mechanism emerges. *Neuron* 16:447–480. (1996)

[5] D. Ferster and K. D. Miller. Neural mechanisms of orientation selectivity in the visual cortex. *Annual Reviews of Neurosci.,* 23:441–471. (2000)

[6] L. Floridi. Is information meaningful data? Forthcoming in Philosophy and Phenomenological Research. (2004)

[7] L. Floridi. Information. In: Mitcham C (ed.). Forthcoming in Encyclopedia of Science, Technology, and Ethics. Macmillan. (2004)

[8] D. Heeger. Optical flow using spatio-temporal filters. *Int. J. Computer Science*, 1(4), 279–302. (1988)

[9] D. H. Hubel and T. N. Wiesel. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol.*, 160:106–154. (1962)

[10] D. H. Hubel and T. N. Wiesel. The functional architecture of the macaque visual cortex. The Ferrier lecture. *Proc. Royal. Soc. B,* 198:1–59. (1977)

[11] S. W. Kuffler. Discharge patterns and functional organization of mammalian retina. *J. Neurophysiol.,* 16:37–68. (1953)

[12] D. Marr and E. C. Hildreth. Theory of edge detection. *Proc. Of the Royal Society of London, B,* 207, 187–217. (1980)

[13] V. B. Mounthcastle. The columnar organization of the neocortex. *Brain,* 120:701–722. (1997)

[14] H. Sompolinsky and R. Shapley New percpectives on the mechanisms for orientation selectivity. *Current Opinion in Neurobiology* 7:514–522. (1997)

ETHICS IN THE COMPUTER PROFESSION

Kim Anttila
Department of Informatics
School of Business, Mälardalen University,
Västerås, Sweden
kaa01001@student.mdh.se

***Abstract*— The computers and the computer science have brought a lot of good to the modern society. But the general public do not always perceive the progress as entirely god. When a decrease in jobs occur because of automation, computerized systems that we rely on goes down and new computer viruses attack, both the general public and the computer professionals realize that we are living in a fragile society. The individual computer professional has some responsibilities but the other parties that are ordering and using the computers and the systems have also responsibilities.**

***Index Terms*— ethics, computer professional, social responsibility, ethical quandaries.**

I. INTRODUCTION

The general public knows through the own experiences and news reports that the development and introduction of computer systems will most likely decrease the number of jobs. That is nothing new; inventions and new technology have since the first days of the wheel made some jobs obsolete. The industrial revolution has, especially since the invention of the spinning machine Spinning Jenny and the steam engine, accelerated the vanishing of certain types of jobs. Few people want to reverse the history to the days before the industrial revolution or pay more for a product or a service but they are at the same time worried about the disappearance of their jobs.

With the introduction of computers there has, besides the problem with the decline in number of jobs, also been an augmentation of other types of ethical problems. Not to say that the computers or the use of them has created new types of ethical questions but the computers have made us more aware of them. The computers have accelerated the impact and the scope of the questions.

To be aware of the consequences there is a need for a discussion of the ethical questions among all involved. Some of the questions are about work ethics, how we behave and work as professionals. Other questions, which probably are even bigger and more important, are about social responsibility since the computer systems ultimately are used in the society. In the early days of technology and also computer science there was not so much consideration either on work ethics or social responsibility. There has after the Second World War been more consideration on the matter and especially after the Nürnberg trials. There it was recognized that the science is not functioning outside the society but within it. For example the medical profession made experiments, even those with lethal outcomes, on concentration camp prisoners. The judges in the trail disapproved that type of actions. The computer industry, though never put on trail, was at the time partly involved in helping the Nazi-regime to commit genocide. It was done by selling and supporting the use of automated machines for sorting of files for national registration that contained information of ethnical race.

I will in this paper discuss the need for ethics, the construction of professional codes and how they can guide the professionals. I will also argue that the professional alone can not take the whole responsibility and that the client who is ordering computer systems needs to consider both the ethical and social consequences. The society itself must also assume responsibility and if there is need, intervene with legislation.

II. WHY DO WE NEED ETHICS?

To answer that question one need to think about the opposite situation. How would a society without ethical consideration be? I assume that no one of us wants to live under the ”state of nature” as Thomas Hobbes described it in Leviathan [1]. In that state every contact with other individuals is a struggle on life or death to obtain the other individuals possessions, or at the best a great effort for acquirement of necessities in life. It would certainly be an unpleasant and socially harsh life, without any ethical considerations and where ”Might is right”.

Mendieta [2] describes the necessity of ethics when writing about the work of the philosopher Dussel saying that the core of Dussel's ethics ”has been the fundamental recognition that human life is vulnerable and fragile, defenseless and abandoned, mangled and suffering. Ethics emerges from our confrontation with this inescapable fact of human existence. It is precisely because we are both incomplete and mangled, suffering and needy that we must be ethical.”

The necessity of ethical consideration and behavior is also apparent in the society when the general public disapproves some discriminating acts of a government agency, the business or even when a person misbehaves on a bus. I think that ethical consideration and moral is like grease in machinery, it lowers the friction among the different parts.

Considering the above-mentioned points makes it obvious that we need ethics also in the computer profession. But which ethical norms should be guiding the professional?

III. DIFFERENT APPROACHES TO ETHICS

In the Western civilizations there are some main lines of ethical theories. The first one is virtue ethics, with roots in ancient Greek philosophy, which is dealing with how individuals should be. A virtuous person does have some desirable character traits that are acquired through education from early childhood and refined in interaction with the society.

Ethics of duties is concerned with the mandatory moral obligations that individuals do have to others. It is also called a deontological theory, from the Greek word ”deon” for duty. The duties that individuals do have to others have varied during the centuries and cultures.

The ethics of rights is based on the freedom to various liberties, freedom from various harms, and the right to help when in need. The UN Universal Declaration of Human Rights has contributed to a clarification on what human rights are.

Utilitarian ethics is about maximizing the pleasure and minimizing the pain. The first version of this ethical theory was requiring a special calculus on the impact of the action to be made, which turned out to be troublesome, laborious, and ambiguous. The version that was elaborated by Mill [3], a Scottish philosopher, is emphasizing the utility for all concerned. He is diminishing the importance of the utility for the actor, which is irrelevant in the consideration. Modern day deceptive implementation of this theory is often, in contrast to Mill's idea, based on the utility for the actor or a smaller group of favored stakeholders.

All of the ethical theories do have drawbacks and benefits but I will not dwell upon those[1]. Most people do have a mixture of ethical principles or private ethical codes in their lives [1]. People do also have different sets of ethical principles depending on the role that the person assumes and those principles could compete with each other [4].

Problems arise often when there are different opinions about ethics. In a project the client could have a short-term cost-benefit approach according to a limited utility model. The computer professional on the other hand could adhere to a professional code of ethics that is based on duty and social responsibility. In a team of developers there are probably also different viewpoints because of upbringing, religion, culture, and class.

The potential differences in ethical standpoints manifest a need for an awareness and discussion on ethics both amongst computer professionals, in dealings with clients, and in the society.

IV. How are the codes constructed?

Ethical codes for a profession are usually developed with the different stakeholders interest in mind [5]. Society should receive due consideration since computer systems have a great impact on it. Employer pays and trusts the professional to perform tasks with a great deal of autonomy. Clients are dependent on the professional's knowledge and judgment and expect a certain level of integrity and professional behavior from the professional. Colleagues are depending on the professional in the teamwork that often is the case and expect also review and help when needed. Organizations where professionals are affiliated claim also loyalty and adherence to the standards of the organization and profession. The profession should also be protected against dabblers since they could harm the reputation [5].

ACM/IEEE-CS Software Engineering Code of Ethics and Professional Practice [6] is an example of professional code of ethics. It is intended to be an international code for an international organization. In the preparation of the code there was taken great precaution in getting opinions from different parts of the world. The ACM/IEEE-CS code is therefore based on principles from ethics of rights and virtue [7]. The ethics of rights was mainly emphasized by the members from US and the virtue ethics by European members [2], which also reflects the possibility of ethical differences in a globalized economy. The authors of ACM/IEEE-CS code acknowledged that no code could be perfect and that the professional has an own duty to be vigilant on new ethical questions.

Many of the codes do have a focus on the work ethical questions, some minimal consideration for social responsibility and a complete absence of reflection on global issues. This is understandable because it is not the aim of the codes.

The codes show that there is an interest from the profession to consider ethical questions and to behave in an ethical manner. They also fulfill their purpose in a sufficient manner in guarding the standards of work ethic.

V. What are the benefits of codes?

Martin and Schinzinger have identified some of the principal purposes of ethical codes [8]. The codes should inspire to an ethical behavior and give general guidance, which could be complemented in other documents. They could in situations where there is a conflict of duties give moral support, have a disciplinary or deterrent effect on individuals that could lapse from ethical behavior. In education of professionals in classrooms or in organizations they could serve as the basis of discussions adding to the mutual understanding. Since the public in general is affected by computer systems the codes could also enhance the image of the profession showing that it is concerned about the ethical behavior of it's members and making self-regulation possible. Once codes are drafted and accepted there will be a mutual understanding in the profession, which gives peace and quiet.

Many of the codes are like the ACM/IEEE-CS mainly dealing with questions of work ethics. If the codes get in to social issues they do it only in a superficial manner giving no guidance on behavior. When Oz [5] examined several codes none of them mentioned anything about global issues. The absence of global issues and low commitment on social issues is understandable since the individual computer professional does have little influence in the subjects.

Professional codes of ethics are a necessity if the occupation wants to be accepted as a respectable profession[9]. If not, the occupation will be in the same league as the horse-dealers. Since the codes are so weak in helping out in questions of social responsibility and they are only intended to be used by individuals there is a need for an other solution to look after those concerns.

[1]The interested reader could consult for example ”Moral Philosophy trough the Ages” by James Fieser

[2]Personal communication with Dr Don Gotterbarn

VI. What could be ethical questions in computer system projects?

The ethical questions could be divided into two different categories, work ethical and social responsibility. Much has been written on the subject of work ethics and computers since Mason wrote a paper on PAPA (privacy, accuracy, property and accessibility) describing some of the ethical key concern in the information age [10]. Other topics in work ethics are for example the lack of user involvement [11], deficiencies in functions in the computer systems, exceed in cost, projects run over the time, how to conduct the research for requirements, and use of models, methods and techniques. If the professional do not have the knowledge of a novel technology, should then the client pay for the education [12]?

In the subject of social responsibility there is the classic question of vanishing jobs through business process reengineering and automation [13]. Visala [14] has composed a list of greater ethical dilemmas that has arisen with the appearance of computers and information systems. This dilemmas are complicated and as Visala notes, a question for the research and planners:
They are the dilemmas of

- efficiency and justice
- a possible contradiction between knowledge and good
- bureaucracy and freedom
- automation and human control
- tradition and innovation
- rational choice with insufficient knowledge
- personal views and social decision making
- opposing interests in organizations
- constraining circumstances and freedom of planning

It could be argued that the ethical questions presented above are not unique for the computer profession and that they are present in other professions also. That is correct but the issue is how they are approached in the computer profession. If the ethical questions are not treated in a systematic way in other professions do not make them less important for the computer profession, two wrong do not make one right.

In the whole area of the computer profession there are a lot of ethical questions and there is probably others that will emerge while the technology and the systems are evolving. It is therefore impossible to make a list of ethical questions and it is better to be prepared to deal with them through a systematical work on different levels.

VII. Why do ethical errors occur in computer system projects?

Bayles has identified four common reasons to ethical misconduct in general which also apply to the computer profession [12]. The first is the age-old greed and it could be directly coupled to other misconduct. It is possible to be greedy as long as there is somebody to exploit. The second is pride that prevents an open communication on the workplace and with clients. The third is a misdirected desire to help others violating trusts or thinking that one knows better or has the rights to do a certain action. The fourth is unawareness of possible ethical questions in computer system projects and the consequences of them.

Jackson [15] adds two more reasons to the list; Lack of motivation and lack of will-power.

An interesting example of ethical misconduct is Andersen LPP. The company conducted a study on business ethics and which factors do have influence on the ethical behavior and based on that study they gave recommendations in the matter [16]. Although Andersen LPP had this deep knowledge in business ethics it was not sufficient to internally behave in an ethical way. FBI investigates the company in the economical scandals concerning Enron.

Some of the reasons for ethical errors are possible to remedy through codes, peer pressure, supervision and discussion in an open mode. Some of them need more effort and time to correct.

VIII. Who is responsible and on which level?

Bayles [12] presents a model containing five different relationships with different levels of responsibility, which could help us answer the question above. In the agency relationship the client dictates what ought to be done and the professional only executes the orders, all the responsibility is on the client. The opposite relationship is the paternalistic where the roles are reversed and where all the responsibility is on the professional. A client professional relationship based on the friendship mode could be viewed as a partnership between near friends. Contractual relationship is based on mutual obligations and rights, which is negotiated and agreed upon by two equals.

The most common mode of relationship is the fiduciary type, which is based on trust between the professional and the client. This trust is founded on earlier shown trust and continuity in the relationship. In an ideal fiduciary relationship the client maintains a considerable part of the responsibility and possibility to make decisions and the professional is acknowledged for his expertise. In the reality there will be some intermixing with the other types of relationship.

A. Supplier company's responsibility

The company and its executives do have a responsibility to orchestrate the different expectation that stakeholders have.

The executives are, according to Barnard [4], not expected to exercise so much leadership as creating an environment of cooperation. This is certainly true in the case of ethical questions that could, as shown above, be disparate and could lead to heated discussion and disagreements both among the employees and between employees and the executives.

A pressure should never be put on an individual to commit a moral compromise, because that could lead to two different negative outcomes. There could be a loss of self-reliance because of paralysis of action, which could result in emotional stress, sense of dissatisfaction and indecisiveness. The other outcome could be a loss of self-esteem, resulting from sense of guilt or discomfort if the individual violates a important private code [4]. Both outcomes could have a negative effect on the employee's performance with negative results for the company.

B. Client company's responsibility

Company's uses often a business ethic based on a cost-benefit model since their main goal is to earn money. But the client needs also to act in a socially responsible way.

Client company's relationship to the computer professional or to the company where the computer professional works, is depending on the clients level of knowledge about computer technology. There will be a relationship based on contract or agency when the client has a good understanding and has been involved in computer system projects earlier. On the other side the relationship will probably be fiduciary if the client are new to computer system projects [12].

The client who orders a computer system does have a great responsibility when choosing projects to realize. They do also choose the contractor and in that deal they are obliged to communicate the values that they stand for.

C. Responsibility of computer professionals

To the individual computer professional it is quite obvious that he has obligations to the employer, client and colleagues, since they are in the locus of consideration due to the proximity.

Employed computer professionals do have a relationship based on fiduciary and some of the aspects of agency since the employer trust the computer professional in many matters but still the employer has authority over the computer professional [12].

As in other circumstances the computer professional has firstly a loyalty towards the employer if he is employed, or to the client if he is hired. He needs to communicate all the concerns and doubts directly to the employer or to the client and not to the users or the employees of the client.

There is a possibility that the computer professionals will through their supply of services in computer systems, by the clients employees be considered as sympathizing with their employers goals to reduce the number of jobs [12]. Therefore he needs to communicate well with the client and employees, if he is permitted by the client, to clarify all the consequences.

In connection to the point above Hirschheim and Klein [17] describes various ways to function as a information system developer that also could be applied on computer professionals.

The first are as systems expert where the computer professional act as a strict engineer, only working with hard measurable facts and he is an expert in technology, tools and methods. The management is the group which decides on profitability and the objectives of the system, and the developed system archives this goals and are the ideal of profit maximization.

The facilitator is the second mode of function as a computer professional where the emphasis is on achieving consensus among all concerned. Computer professionals assuming this function works also as change agents, trying to help the different parties in the project to find their preferred views of the system. The implementation of the new system should not disturb the social equilibrium.

The third in Hirschheim and Klein's list is the labor partisan. The computer professional could side with the management and their interest resulting in a system where there is a loss of jobs, deskilling of jobs and so forth. At the other hand the computer professional could side with the laborers and pursue their goals. The system developed would be a system that enhances the labor's skill and making the job more interesting with a better product as a result. There is much consideration on the value of work and the situation of the workers.

The fourth function is as emancipator or social therapist and is hypothetical unlike the other three. The systems development and use is laid on the foundation of rational discourse. System developers using this approach tries to understand and develop a system that emancipates suppressed interests and works ”toward a state of justice, freedom, and material well-being for all” as Hirschheim and Klein describes it.

None of the functions described above, neither the classical nor the more novel, do consider social responsibility in any higher degree. There is considerations on the level of a specific workplace and the proximity of it, but nothing on the society or global sphere. Few if any of the existing development models or methods do take the ethical questions in consideration.

It is in many cases unfair to require that the computer professional should know about all the consequences of a computer system on social and job level. To estimate the social consequences of a computer system under development the computer professional need to know a lot of sociology and to estimate the consequences on the jobs they have to have additional knowledge. The individual computer professional is the weakest and most vulnerable party, besides workers affected, in the organizational structure.

D. Responsibility of the society

Further on, the question is also who is making the decisions in questions of social responsibility: should only the free market do it or do the society have a duty to intervene? Some of the questions about social impact of computer systems are political and remains outside the control of the client, supplier and computer professional. So the question is, which kind of society do we want to have? How much should we automatize and put under control of computers? What values do we appreciate?

IX. Conclusion

The conclusion of my paper is that there are a lot of questions to consider at different levels. Computer professionals do not work in a vacuum and there is a need to go beyond the professional codes of ethics that are based on work ethics. Many of the ethical questions are impossible for the individual computer professional to take responsibility for. The profession, employers, and clients needs also to reflect on the ethical questions and take their part in showing social responsibility.

Hitherto there has been a progress in the development of the more or less hard components, models, methods and techniques, in the computer system development. The individual has been forced in to the use of these technicalities and the focus has been on the compliance. There has been little or no consideration on the development of the individual's character or the companies' ethical environment, which could motivate to better compliance.

The work on ethics in computer science has until today dealt with the micro level of ethical questions and because of that there has been progress in the writing and discussion of professional ethical codes. It is probably time to start the discussion and work with the next level, the level of the the companies that employ and develop computer systems, even called the meso level. This should not only be done with the help of corporate ethical codes, but with a systematical approach where the ethical questions are openly discussed.

References

[1] J. Fieser. *Moral Philosophy trough the Ages*. Mayfield Publishing Company, Mountain View, CA, 1st edition, 2001.

[2] E. Mendieta. Review essay: Ethics for an age of globalization and exclusion. *Philosphy & Social Critisism*, 25(2):115121, 1999.

[3] J. S. Mill. *Utilitarianism*. Daidalos, Uddevalla, 2003.

[4] C. I. Barnard. *The Fucntions of the Executive*. Harvard Univ. Press, Cambridge, Mass, 30 edition, (1938)1968.

[5] E. Oz. Ethical standards for information systems professional: A case for unified code. *MIS Quarterly*, pages 423–433, December 1992.

[6] ACM/IEEE-CS joint task force on Software Engineering Ethics and Professional Practices. Software engineering code of ethics and professional practice. *http://www.acm.org/serving/se/code.htm*, 2004-04-10.

[7] D. Gotterbarn. Two approaches to computer ethics. *SIGCSE Bulletin*, 31(2):11–12, 1999.

[8] M. Martin and R. Schinzinger. Codes of ethics. In D. H. Johnson and H. Nissenbaum, editors, *Computers, Ethics & Social Values*, pages 576–580. Prentice Hall, Englewood Cliffs, N.J., 1995.

[9] N. E. Ford, G. Gibbs. A mature profession of software engineering. Technical Report CMU/SEI-96-TR-004, ESC-TR-96-004, Software Engineering Institute, Carnegie Mellon University, Pen, jan 1996.

[10] R. O. Mason. Four ethical issues of the information age. *MIS Quarterly*, 10(1):5–13, 1986.

[11] The Standish Group. Unfinished voyages. *http://www.standishgroup.com /sample_research/unfinished_voyages_1.php*, 2004-05-10.

[12] M. D. Bayles. *Professional Ethics*. Wadsworth Publishing Company, Belmont, CA, 2nd edition, 1989.

[13] K. C. Laudon and J. P. Laudon. *Management Information Systems: Organization and Technology*. Prentice-Hall, Upper Saddle River, N.J., 3rd edition, 1999.

[14] Seppo Visala. Ethical dilemmas in information systems research and development. In *Proceedings of the 1992 ACM SIGCPR conference on Computer personnel research*, pages 210–220. ACM Press, 1992.

[15] J. Jackson. *An Introduction to Business Ethics*. Blackwell Publishers, Oxford, 1st edition, 1996.

[16] J. S. McClenahen. Your employees know better. *Industry Week*, 5(248):12–13, March 1999.

[17] R. Hirschheim and H. K. Klein. Four paradigms of information systems development. *Communications of the ACM*, 32(10):1199–1216, oct 1989.

Gender Distribution Normalization in the Computer Game Development Arena

Mikael Sandberg
Department of Computer Science, Mälardalen University
mikael.sandberg@mdh.se

***Abstract*— Gender bias in the computer game arena is far from new or, what we believe is even worse, has come to be the norm of today. We see more games targeted towards males of the ages 13–25 developed today, then ever before.**

Computer games are also becoming more violent at an alarming rate. A few computer games could even be considered to be the equivalent of interactive *SPLATTER-movies*[1].

Some computer games promote unethical behavior that most likely reinforce anti-social behavior and might even encourage actual behavior changes in players.

We believe that an increase in female participation in game design, development and not least, game playing, might be one of the few viable solutions in decreasing the violence trend in the computer gaming arena.

I. Introduction

Games play a big roll in the human learning and development process, both mentally, intellectually and socially. Playing games involve the voluntary activity that is intrinsically motivating, have some kind of activity (often physical), and may incorporate make-believe qualities[1]

The computer gaming industry prey on this human need, to play games, and are popularly believed today, to have a gross turnover exceeding even the mighty movie/music industry.

We will not try to define what is a good or bad computer game in a gender perspective. We will however give a partial account for what others believe to be good and bad in a gender perspective. We will also try to account for our claim that an increase of female participation in computer game, design, development and gaming would have an positive impact on decreasing computer game violence and unethical behavior.

This article will not focus on the economical aspects of the problem. But we can not refrain from stating a few facts on the subject.

Sheri Graner Ray states in her book[2] that 51% of the women in the United States of America use Internet on a regular bases, and that 70% of the female internet users occasionally play online games.

T.L Taylor state in Convergence[3] that $20\% - 30\%$ of the players are women in the three largest online games, Archeron´s call, Ultima Online and EverQuest.

Shiels[4] cite Clarrinda Merripen, Cyberlore Studios, who stress: “Females comprise 80% of the consumer dollars and very few of them are buying games on an ongoing basis”.

This however does not seem to be a big enough incentive for the computer gaming industry to invest in the development of games that could be construed as women friendly computer games. We believe it is due to the fact that there is another factor the gaming industry can not overcome, a cultural change, that a computer is not only a working tool but also an entertainment unit rivalling or even surpassing the entertainment value of the television set We believe this will change rapidly and within a period of 2–5 years will no longer pose a problem for the gaming industry.

A. Games and Violence

Historically games in general have some element that contain violence. Even games that are now considered classical, like Chess, Go[2] and Awari[3] contain elements of violence like, capturing opponents pieces and removing (killing) pieces from the board etc.

Society have accepted this violent element in the games and take them for granted today. This *Accepted Game Violence* (*AGV*) has been part of our society for generations and pose no ethical problems, as opposed to *Incorrect Game Violence* (*IGV*) that violates most of our ethical believes and standards.

Some traditional games played, mostly by young children, are borderline and could be placed in the middle

[1] A SPLATTER-movie is a movie genre that contains excessive amounts of violence and use extreme amounts of artificial blood in violent scenes hence the name.

[2] Go is an ancient Chinese strategy board game.

[3] Awari is an old African game of the Mancala family of board games.

between *AGV* and *IGV*. In this group of games are among others: “Cowboys and Indians”, war enactment (WW I, WW II, Vietnam, Desert Storm, and lately Iraq War). Depending on the ethical frame some of the games might pose an ethical problem and some do not, but still most consider them harmless.

B. Are Computer Games any different?

Computer Games are very similar to traditional games in many ways. One major difference is however that players do not have to use their imagination to the same extent as is found in traditional games. Computer games provide a more rehearsed setting and often hand a preset stage to the players to play in, giving the players a limited use of their imagination with in that stage.

The multitude of different types of computer games are as diverse as traditional games and range from educational to leisure.

Computer game violence cover the whole spectrum, from none violent and *AGV* to *IGV*. The trend is however towards more games with *IGV* then none violent or *AGV* computer games.

Computer games of tomorrow would most likely include:

- Better realism.
- Multi-play (e.g. player vs. player).
- Continuous world[4]

II. Games and the Computer Gaming Industry

The computer gaming industry is responding to the players demand of violent computer games[5], the computer games become more graphic and resemble reality to an uncanny degree. We have come to the point where computer game violence with most certainty, is having a negative effect on players. We believe that this increase in demand of violence is, on the most part, due to the fact that the audience of computer games are made up of males of ages 13–25. We see less violence in games with a higher fraction of female players. We also believe that the computer gaming industry might interpret the demand for increased violence as a major selling point to a degree that does not correspond to the actual computer game players demands.

Governments have already banned a few particularly cruel and unethical computer games. More banns will most surely follow, since the trend is towards increasing violence and unethical aspects in computer games rather then to promote ethical and non violent behavior.

[4]In a multi-player environment the world in which the game take place continues even if a player leave the game

Computer Game Developers are arguing that developing games, their version of good games, should not be subject to any legislation and doing so would be an impairment of their freedom to express their artistic creations(“Freedom of speech”).

A government banning system, national or international will most likely have a profound negative effect on the gaming industry and game players. This would most surely result in decreased revenues and what might be considerably more difficult to restore, the the public opinion against computer games.

When will parents, school organizations and society in general take action against the extremely violent and unethical computer games? We believe it will be soon. Computer game violence is getting more media coverage now then ever. A computer game (DOOM[5]) was even an issue in the columbine trial, however; the game was not considered by the court to have influenced the behaviour of the boys that committed the crime to such a degree that it could have influenced their actions in any major way. We do however believe that this might raise a debate on the subject.

III. Computer Games and Gender

At the same time as the violence and unethical aspects increase in computer games we see a decline in female players on those computer games, but an increase in female players in general.

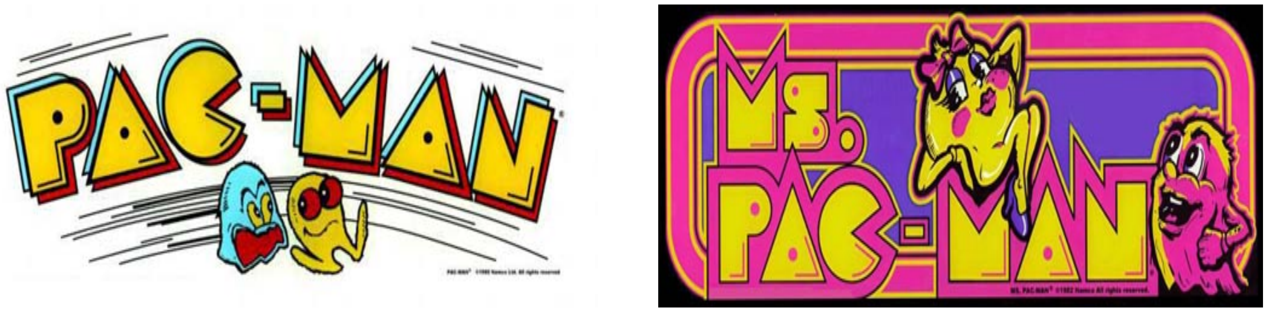
Male and female players both like to play computer games and both genders, in general, enjoy the same aspects of games but there are slight differences in the priority of these aspects. For instance, equal sum games[6] are less preferred by female players then male players[2].

It is popularly believed that one of the major aspects of the female player shortage is due to the violent nature of most modern computer games. This is not our belief and we concur with Sheri Graner Ray, who in her book concludes that female players dislike unmotivated violence to a much higher degree then male players[2] but they do accept motivated violence *(AGV)* to the same degree as male players. This is in our view, one of the major reasons why a gender neutral game would contain less violence and definitely less *IGV*.

Female players do however feel locked out of the game arena due to the fact that it is so male dominated[4]. This does include both the development of computer games as well as the playing of computer games.

[5]DOOM was one of the first: First Person Shooters, created by ID Software and was played by many players at the time

[6]Equal sum games are games that one side have to loose something for the other side to gain something.



(a) Pac-Man

(b) Ms. Pac-Man

Fig. 1. Pictures showing the head mount of both the original Atari Pac-Man arcade game and the head mount of follow up Ms. Pac-Man arcade game.

It is very popular to try to find gender specific stereotype patterns[6]. It is believed that finding such a pattern would increase the chance of developing a game that would appeal to women. This stereotype pattern however, seems to fit both genders to an equal degree and are therefore hard to construct and could be considered as gender neutral stereotyping for the most part rather then gender specific.

A. Gender Neutral Games

We believe that a computer game could be enjoyable for both genders and that games should be gender neutral as opposed to gender specific. If games are gender specific they would promote rather than neutralize the gender segregated computer gaming arena. We strongly believe that this segregated arena is one of the biggest hurdles and that it is the most important one to overcome.

This hurdle includes the fact that computers are intimate machines[7] (e.q they can be perceived as logical or "thinking"), and the male and female approach towards these intimate machines are a bit different, but are becoming more gender neutral today.

Female players are more tactile and communicative towards solving problems and working with computers. Male players are more likely to take risks and see computer problems as challenges to be overcome, in other words male players work against computers rather then work with them as female players do[7]. We believe that this is changing rapidly and that most see computers as both tools and entertainment machines.

B. Painting it Pink

A popular way of constructing games, that is believed to be appealing to female players, is to "Paint it Pink". A good example of this is the famous computer arcade game called Pac Man (Figure 1(a)). Pac Man was such a success among male players that the company that developed it wanted to do a a computer game that also appealed to female players. They constructed a new version of the same game and named it, Ms Pac Man (Figure 1(b)). The only difference was the pink bow tie and lipstick the avatar was fitted with. Needless to say, it failed to attract the female players, and it only gained a moderate success among the male players as a sequel game to Pac Man.

Painting it pink does not suffice[8], for games to be appealing to both male and female players, developers have to learn what female players want. Developers know what male players want, or think they know at least. We disagree that to accomplish this the developer has to be female[4].

However male developers might have to relearn and gain new experience that would suit both genders since most of the computer game development is done in a mono gender culture.

IV. CONCLUSIONS

We believe and hope that one day we will look back on this violent and unethical computer game era as the "dark ages" of computer games and marvel over how bad it's games were.

Male and female computer game players are more alike then first meets the eye. Creating games that is appealing to both genders is not such a "big deal". There are plenty of examples in the book industry that contradict the need for gender specific computer games.

We consider the increase of female involvement in the whole computer gaming chain of development and paying of games, as one of the few viable solutions to the increasing number of ethically incorrect games. Producing games that fit both genders do however not exclude all violence, as we stated, games do, more then not, include some kind of violent element.

We think a gender neutral computer gaming arena is very important, since playing games is so important to the development process of children. We think it is better then a gender specific computer games arena. We do nevertheless believe that to some extend that, gender specific games could increase the fraction of female players initially. Those games however should not be “Paint it Pink” versions of games of known games that appeal to male players, but new and innovative computer games that in the best case can be appealing to both genders.

Lastly we would like to cite a book[9] where the following statements nicely summarize what we believe are the right way to proceed developing computer games that is more gender neutral:

- “Girls need to be able to play games where Barbie gets to kick some butt”
- “Boys may need to play in secret gardens or toy towns just as much as girls need to explore adventure islands”

References

[1] L. Rieber, “Seriously considering play: Designing interactive larning environmets based on blending of microworlds, simulations, and games,” in *Educational Technology Research and Development*, vol. 44(2). Association for Educational Communications and Technology, 1996, pp. 43–58.

[2] S. G. Ray, *Gender Inclusive Game Design - Expanding the Market*. 10 Downer Avenue, Hingham MA, 02043: Charles River Media, 2004.

[3] T. Taylor, “Multiple pleasures: Women and online gaming,” in *Journal of Research into New Media Technologies*, vol. 9. Convergence, 2003, pp. 21–41.

[4] M. Shiels, “Games makers miss feminine touch,” *BBC News*, March 2003, available online at: *http://news.bbc.co.uk/2/hi/business/2839019.stm*.

[5] IGN, “Nintendo speaks out,” October 2002, available online at: *http://cube.ign.com/articles/376/376063p1.html*.

[6] J. Durchin, “Girls software introduction: Keys to creating great girl software,” in *Game Developers´ Conference*. International Game Developers Association, March 2000.

[7] S. Turkle, “Computational reticence: Why woman fear the intimate machine,” in *Technology and Women’s Voices*, C. Kramarae, Ed. New York: Pergamon Press, 1986.

[8] C. Now, “Girls and gaming: Gender and video game marketing,” 2000, children Now, Available online at: *http://www.childrennow.org/media/medianow/mnwinter2001.htm*.

[9] J. C. et al, “Complete freedom of movement:video games as gendered play spaces,” in *From Barbie to Mortal Kombat: Gender and Computer Games*, H. Jenkins and J. Cassell, Eds., November 1998.

Is Big Brother a human necessity?

Omar Bagdadi
Swedish National Road and Transport Research Institute
omar.bagdadi@vti.se

Abstract

Is the fact that governments and companies uses the information technology to watch over and register peoples living habits an immoral act and hence must cease or do we in fact need a "Big brother" watching over us and can the invasion of personal privacy be justified? According to consequence based theory, i.e. Utilitarianism, the outcome determines if an action is morally justified meaning that action is right if it promotes the maximum good for everyone, or at least greater good for the greater number. This also implies that the ends sometimes are more important than the means. How do we value our own privacy and should we in fact encourage government to use any means necessary to keep us safe from criminals and terrorists?

1 Introduction

Information is power. The value of personal information, especially information about commercial purchases and preferences, has long been recognized. It did not take long before companies came to realize that such data is very valuable. The expense of marketing efforts gives businesses a strong incentive to know as much about consumers as possible so they can focus on the most likely new customers. Today, any consumer activity that is not being tracked and recorded is increasingly being viewed by businesses as a waste of money.

On the Internet, where every mouse click can be recorded, the tracking and profiling of consumers is even more widely spread. Web sites can not only track what consumers buy, but what they look at, for how long, and in what order.

With the end of the Dot Com era, personal information has become an even more precious source of hard cash. And, not to forget, we use the Internet not just as a shopping mall, but to research topics of interest, debate political issues, seek support for personal problems, and many other purposes that can generate deeply private information about their thoughts, interests, lifestyles, habits, and activities.

The development of information technologies continue to accelerate. The technology is becoming more powerful, cheaper and easier to manage, which makes it available to more people and organisations than ever. At the same time it is getting easier to take part of and share digital information. The explosion of computers, cameras, sensors, wireless communication, GPS, biometrics, and other technologies in just the last 10 years is creating a surveillance society. Scarcely a month goes by in which we don't read about some new technology that can be used to invade people's privacy, from face recognition to implantable microchips, data-mining, DNA chips, and even "brain wave fingerprinting." I believe that there are no longer any technical barriers to the Big Brother regime portrayed by George Orwell.

As the use of information systems has become pervasive in advanced economies and societies at large, several ethical and social issues have moved into the forefront. The most important are issues of individual privacy, property rights, universal access and free speech, information accuracy, and quality of life [1].

The information society makes it particularly important to explore the perspectives of privacy as well as the interests and values of stakeholders. The possibility to speed up information exchange and to aggregate information may substantially alter our previous understanding of privacy interests. As mentioned earlier, there is an element of power associated with the interests and values of stakeholders. Certain stakeholders are in a better place to serve their interests and satisfy their internal values than others. This asymmetry of power may raise risks for the privacy of the weaker stakeholders.

Curiously, at the same time as privacy concerns, in particular, privacy over the Internet, is discussed, people tend to have no problem revealing personal information during for instance online shopping. This is often done because of convenience, discounts and other incentives, or perhaps a lack of understanding the consequences [1].

The greatest threat to privacy comes from the business of recording and collecting everyday transactions such as supermarket records over every item being bought by customers using a discount card, any use of a bank or credit card or telephone calls [14].

Telecom operators are installing equipment that allows them to eavesdrop on a conversation or even track the location of anyone using a mobile phone.

Although most people know that information is collected, they often don't have a clue how much or exactly what information that is stored or how it will be used. One single company in Arkansas, USA, Acxiom Corporation has a database combining public and personal information that covers 95% of the American households, and made $958 million in revenue last year from selling people's names, addresses and profiles.

The computer ethicists have expressed concern about privacy and fear that the information gathering and exchange can easily turn the "information society" into "surveillance society" [2].

We must not forget that the rapid growth in information technology and the emergence of the Internet and World Wide Web over the last several years have been beneficial to both individuals as well as societies. For instance it:

- Enables people in Third World countries and rural areas to enjoy products and services which otherwise are not available to them
- Facilitates the delivery of public services at a reduced cost, increases effectiveness, and even improves quality
- Makes it possible for more individuals to work at home, and to do less travelling for shopping, resulting in less traffic on the roads, and lower air pollution
- Online shopping allows some merchandise to be sold at lower prices since the "store owner" may not need a physical place and full inventory

Is it possible that we, in our effort to become more and more efficient and adding what we believe is more values to our lives by using the full potential of our networked world, without realizing it created an invisible, all seeing, electronic panopticon, a world where we are continuously under surveillance?

2 The panoptic effect

Jeremy Bentham, the British philosopher and social reformer, published his plan for the Panopticon penitentiary in 1791. Essentially, it was for a building on a semi-circular pattern with an 'inspection lodge' at the centre and cells around the perimeter. Prisoners, who in the original plan would be in individual cells, were open to the gaze of the guards, or 'inspectors', but the same was not true of the view the other way.

The key principle was inspection, through inspection of a specific kind. The more constantly the persons to be inspected are under the eyes of the persons who should inspect them the more perfectly will the purpose of the establishment be attained. And if such constant supervision proves impossible, prisoners should be given the impression that (s)he is under constant supervision [3].

The effect of the constant supervision is based on true belief of the supervised person. (S)he believes that (s)he is under constant supervision and hence acts thereafter.

This phenomena has been used as early as during the middle-ages when the civil authority were in the hands of the only educated class, the churchmen who in their teaching of what was right and wrong painted out God as the "all-seeing" and "all-knowing" judge with the power to punish even beyond the grave. Thus believing in God and the power of God created a stimulus, though not a philosophical justification, to behave according to the divine wishes.

Today this panoptic effect is realized through information systems. Intelligent systems are able to scan and identify individuals from video images and due to the already low prices and small sizes of cameras and other electronic surveillance tools, camera monitoring should be possible almost everywhere not just in airports or other areas that are normally associated with a high level of security but also in other public areas.

A new type of surveillance that is becoming possible is the collection of information about an identifiable individual, often from multiple sources, that can be assembled into a portrait of that person's activities. Many computers are programmed to automatically store and track usage data, and the spread of computer chips in our daily lives means that more and more of our activities leave behind "data trails." It will soon be possible to combine information from different sources to recreate an individual's activities with such detail that it becomes no different from being followed around all day by a detective with a video camera.

3 The concept of Privacy

Most people would agree that privacy is a civil right, and that we should not be deprived of this civil right for other than legitimate reasons. The difficulty lies instead in how to decide where the legislative boundary should be drawn for the right to privacy, in order to protect the individual from an undue intrusion of his/her privacy.

However, privacy is not a straightforward concept. It can be interpreted from many different perspectives. One reason for this is that privacy is a relational and a relative concept. Often, there is a thin line between the need to disclose information for the benefit of some individuals and the need to safeguard the privacy of some individuals by not disclosing this information.

Although privacy has been discussed since Aristotle there's still not a single definition of privacy. [10]. Two definitions listed in the Oxford English Dictionary are:

a) privacy is the state or condition of being alone, undisturbed, or free from public attention, as a matter of choice or right.
b) privacy is absence or avoidance of publicity or display, a condition approaching to secrecy or concealment.

Both of these definitions are valid and originate from early 17th and late 14th century respectively but are limited to physical privacy, i.e. the right to be left alone. This can also be seen as the most basic definition of privacy. A newer and richer definition stated by Warren S. and Brandeis L. in 1890 is that privacy can also be understood as having control over information about oneself. More precisely, having control over how much and what information that is spread and to whom. [10].

While personal privacy is considered important in Western democratic societies it has not a global value, meaning that different cultures value privacy in different ways. While most countries has adopted a minimal conception of privacy that is more or less common, the differences in customs have formed the rich conception of privacy and determines how, for instance, personal information about someone should be used or shared and in which situations some restrictions of areas or information are in place.[11]

A country where personal privacy in the context of being alone and free from public display is almost impossible to maintain is in Japan, mainly because of its very high population density. In order to maintain a "virtual" private sphere the Japanese has created an "as if" convention, similar to the one bartenders or buss drivers has, meaning that even if someone may gain information about someone (s)he should act "as if" (s)he has not.

It is also important to remember that in order to understand how privacy is valued in a country and what information that is to be considered as private one must understand the structure of the society.

3.1 Privacy before and after 11 September 2001

In the last decade, there has been a great deal of public debate on what measures to take to protect individual "privacy", mainly due to the increased use of the Internet. But more recently, specifically after September 11 2001, there has been a similarly intense debate about how to protect "security."

This greater attention to security has created a general sense that privacy has become a less important issue. We can sometimes see headlines like "security vs. privacy," where it is stated that there is a trade-off between security and privacy. Greater security can often be accomplished through increased information for the security forces; this means more surveillance in public places, such as airports and railway stations, but also at sports arenas and shopping centres. This also means greater information gathering, and information sharing. All these actions can be argued as necessary security measures but it also raises the risk of privacy erosion.

As one sign of the changed times the Bush Administration proposed new legislation, the USA-

PATRIOT Act [9], less than a week after the attacks. In the area of wire-taps and electronic surveillance, the proposal contained a number of provisions that had been previously rejected by Congress as too pro-surveillance. It included other new surveillance powers that had not ever been subject to any hearing or debate in USA congress [12].

Noticeable is that the year before, the Clinton Administration had proposed updating the same laws in ways that also updated law enforcement authorities while being more protective of privacy. The House Judiciary Committee, with an overwhelming bipartisan majority, had also made some changes to the bill to substantially increase the privacy protection.

Now, following the attacks, the previous legislative that were aiming toward greater privacy protections suddenly shifted to greater government surveillance powers than anyone would have seriously proposed only a year earlier. The USA-PATRIOT Act passed, in a new record time, on October 25, 2001 without the hearings that normally are being held before ruling on controversial bills. Critics of the Act were able to make few amendments during its rushed consideration, although some of the most worrisome surveillance provisions will sunset in 2004 [12]

The surveillance provisions of the USA-PATRIOT Act generally illustrate “security vs. privacy.” To take only a few examples [9], the Act:

- Increases the scope of roving wiretaps, where law enforcement can access communications from any device used by a suspect, rather than needing to get a new order for each phone or computer.

- Broadly increases the scope of emergency orders to trace communications, which apply be-fore a judge approves a court order.

- Allows one court order for tracing communications to apply nationwide, rather than requiring a new order in the district where a communications provider operates.

- Allows a much broader category of cases to use information developed under the Foreign Intelligence Surveillance Act, where those subject to wiretaps are not informed of the surveillance even after the fact.

- Permits information developed by a grand jury in a law enforcement proceeding to be shared with intelligence agencies.

- In a “computer trespasser” provision that was never the subject of a Congressional hearing, permits law enforcement officials to set up extended residence at a communications provider to survey the communications of unauthorized users.

4 Ethics

4.1 Introduction

Ethical theory and moral practice refer to human behaviour. In order to act morally one must be able to reflect on his/her behaviour from an ethical point of view, and for this one needs information [7].

Plato stated that people would always act in the right way if they knew what it was: the problem as he saw it was that most people would, could, never attain true knowledge because of their innate limitations. They were destined to live perpetually in a shadowy world of error and trivial amusement. He felt that these people, the majority, should have their lives controlled by the few who were capable of attaining true knowledge through education and reflection: the “Guardians”. True knowledge and ethical correctness were one and the same – to know the right, is to do right. Education was the key and the corrupting influences of popular drama, poetry, and personal property would have no place in his idealised society [4].

The field of ethics can be divided into three branches, metaethics, normative ethics and applied ethics. Metaethics talks about the nature of ethics and moral reasoning. Discussions about whether ethics is relative and whether we always act from self-interest are examples of metaethical discussions. Examples of metaethical questions include:

- What does it mean to say something is "good"?
- How, if at all, do we know what is right and wrong?
- How do moral attitudes motivate action?
- Are there objective values?

Normative ethics is interested in determining the content of our moral behaviour and seek to provide action-guides, procedures for answering the question

- "What actions are good and bad?"
- "What should we do?

Applied ethics attempts to deal with specific realms of human action and to craft criteria for discussing issues that might arise within those realms. The contemporary field of Applied Ethics arouse in the late 1960s and early 1970s. Today, it is a thriving part of the field of ethics. Numerous books and web-sites are devoted to topics such as Business Ethics, Computer Ethics, and Engineering Ethics. [13]

Of these three branches of ethics it is normative ethics that is going to be discussed further in this paper.

4.2 Normative ethics

As mentioned, normative ethics focus on the consequences which any action might have. Thus, in order to make correct moral choices, we have to have some understanding of what will result from our choices. When we make choices which result in the correct consequences, then we are acting morally; when we make choices which result in the incorrect consequences, then we are acting immorally.

The idea that the moral worth of an action is determined by the consequences of that action is often labelled consequentialism. Usually, the "correct consequences" are those which are most beneficial to humanity - they may promote human happiness, human pleasure, human satisfaction, human survival or simply the general welfare of all humans. Whatever the consequences are, it is believed that those consequences are intrinsically good and valuable, and that is why actions which lead to those consequences are moral while actions which lead away from them are immoral.

One of the normative ethical theories that has been widely spread the last two centuries and has played a significant role in law, politics and economics is Utilitarianism.

In its political philosophy Utilitarianism bases the authority of government and the sanctity of individual rights upon their utility, thus providing an alternative to theories of natural law, natural rights, or social contract.

With different factual assumptions, however, Utilitarian arguments can lead to different conclusions. If the inquirer assumes that a strong government is required to check man's basically selfish interests and that any change may threaten the stability of the political order, (s)he may be led by Utilitarian arguments to an authoritarian or conservative position.

4.2.1 Is privacy intrusion justifiable?

A Utilitarian considers that privacy intrusion is wrong since it affects a person's prosperity in a negative way and such an act should therefore be punishable to prevent it from happening again. [6]

At the same time some level of privacy intrusion may be justified in order to increase the level of security and for the prevention of serious crimes. It must be added that the intrusion should be carried out in secret and never be revealed. Is this double standard acceptable?

In 2001 the law enforcements in Tampa used a biometric system that relied on facial recognition at Super Bowl XXXV [5]. The system consisted of cameras surreptitiously scanned spectators' faces to capture images. Each image was than processed through algorithms that measured facial features from these images. For example, the distances and angles between specific geometric points on the face like the mouth extremities, nostrils, and eye corners. This was done to produce a "face print." This face print was then instantly searched against a computerized database of suspected terrorists and known criminals to recognize a specific individual. A match would have alerted police to the presence of a potential threat.
One of the questions that rises is whether this is yet another privacy intrusive system or a standard identification technique no different than the normal watching that security personnel normally performs.
One could argue from a Utilitarian point of view that although facial recognition systems are privacy intrusive it might be the right thing to do anyway as long as it managed without anyone knowing about it.
The real issue comes if the system creates a false alarm, causing someone to be wrongly accused of being, for example, a member of a known terrorist group. As one could imagine, this erroneous information could not only cause a false arrest of a person but also end up inside one of NSA's (National Security Agency) databases of known terrorists.

Another question that arises is if it, according to Utilitarianism, is justified to invade someone's privacy for security reasons. Is it then also possible to justify the any kind of privacy invasion regardless of the cost?

As mentioned earlier, Utilitarianism strives to achieve "the greatest happiness for the greatest number" [8], meaning that the benefit must be higher than the cost.

The benefit of using facial recognition technology in Tampa was increased level of security during the Super bowl event. But what was the cost and was the benefit higher?

5 Discussion and Conclusion

So if we have privacy, we have the ability to control how much information about ourselves is revealed or withheld in our relationships with individuals and with organizations. If we lose that control, then it would seem to follow that we lose our privacy.

If an organization has gathered information to be recorded in that organization's computer database, and if this information is not accessed by other organizations or exchanged with other computer databases without that individual's consent, then it would seem that no breach of that individual's privacy has occurred. No breach of privacy has occurred because the individual, in voluntarily giving information to the organization, has not in fact lost control over that information. If, however, that information is subsequently exchanged -- e.g., matched or merged – with information in other databases, without the consent of the individual who initially granted the information, then that individual has lost control over information about himself/herself.

There is a sense in which computer matching closely resembles the kind of monitoring and surveillance procedures that authorities used in the above example. Although systematic searches of personal computer records may not seem to be as intrusive as physically intercepting personal mail, the result is the same. Both techniques, in their efforts to track down potential law violators, are incompatible with personal privacy.

It would seem however, that many find the kind of electronic surveillance implicit in computer matching less intrusive, or at least less objectionable, than the physical surveillance and monitoring techniques used for example intercepting mail from co-workers within corporate business

A common line of reasoning that is frequently offered to defend a practice like computer matching and video surveillance is that privacy is a right, and rights are not absolute so when a person commits a crime, (s)he forfeits his/her right to privacy.

Even if this kind of reasoning assumes that we have an explicit legal right to privacy and that all legal rights are or ought to be conditional only, certain problems with this line of reasoning still remain.

The problem lies in the fact that it was in the act of matching records of several innocent individuals that a "hit," identifying on or more alleged criminals, could be generated. So even if criminals do forfeit their right to privacy, it would seem to follow that in the process of determining who many of these criminals are, several innocent individuals will be required to forfeit that right as well.

How should the controversy be resolved?

We have seen that the controversy arises because we believe we have a right to privacy and an obligation to protect that right, and we believe we have an obligation to support law-enforcement agencies in their use of information technology to identify and track down potential violators of the law. If we wish to comply with our obligation to preserve personal privacy, and if we accept the view that having privacy is having control over information about oneself, then we cannot consistently support the practice of matching computer records to track down individuals. Conversely, if we comply with our obligation to support law enforcement agencies in tracking down individuals, then we cannot protect our individual privacy.

Initially it might seem to be in our best interests to support the use of computer matching to identify potential violators of the law. If we approve this practice, however, then it would seem to follow that we must not really value our own privacy as much as we may originally have thought. Ultimately, we must decide whether we really value personal privacy and want to protect it or whether to use our information technology to its fullest potential. We may decide that the benefits of matching computer records to apprehend criminals outweigh any loss of personal privacy. Utilitarian arguments could no doubt be advanced to defend that position. Here the end achieved in making our society a safer place could be the reason given to justify the means of using a technique that is incompatible with personal privacy. While many of us would agree that such an end is desirable, we must also understand that in supporting the practice of matching, we will allow ourselves to be used as the means to achieve that end.

6 References

[1] Priscilla Reagan, “Privacy and Commercial Use of Personal Data: Policy Developments in the United States” *Journal of contingencies and crisis management, Volume 11, Number 1*, Blackwell Publishing Ltd, Oxford, UK, 2003, p.1.

[2] Deborah G. Johnson,”The Blackwell guide to the philosophy of computing and information / edited by Luciano Floridi”, Blackwell Publishing Ltd, USA, 2004, pp. 65-75.

[3] David Lyon, “The Electronic Eye: The Rise of Surveillance Society”, University of Minnesota Press, 1994, pp. 57-80.

[4] Chritopher E. Hackley, “Ethical Perspectives on the Post modern Communications Leviathan”, Journal of Business Ethics 20, 1999, pp. 15-26.

[5] John D. Woodward Jr., “Super Bowl Surveillance, Facing Up to Biometrics”, Rand Arroyo Center, 2001.

[6] Torbjörn Tännsjö, “Grundbok i normativ etik”, Stockholm, 2000, pp. 25-46.

[7] Berndt Carsten, ”Minds & machines #14”, Stockholm, 2004, pp. 67-80.

[8] Michael Parkin, “Microeconomics, sixth edition”, Pearson education Inc., New York, 2003, pp. 101-120.

[9] Uniting and Strengthening America by Providing Appropriate Tools Required to Intercept and Obstruct Terrorism (USA-PATRIOT Act) Act of 2001. http://www.epic.org/privacy/terrorism/hr3162.html

[10] Judith DeCew, "Privacy", The Stanford Encyclopedia of Philosophy (Summer 2002 Edition), Edward N. Zalta (ed.), URL = http://plato.stanford.edu/archives/sum2002/entries/privacy/

[11] Masahiko Mizutani , James Dorsey and James H. Moor, “The internet and Japanese conception of privacy”, Ethics and Information Technology, Kluwer Academic Publishers, Netherlands, 2004, pp.1-8.

[12] Anders R. Olsson, “Efter 11 september 2001: - Kan Storebror hejdas?”, VINNOVA, 2003.

[13] James Fieser, ”Ethics”, Internet Encyclopedia of Philosophy, http://www.iep.utm.edu/, mars 2004.

[14] “The Surveillance society”, Economist Vol. 351 Issue 8117, Economist Newspaper Limited; London, 1999.

PRIVACY OF COMPUTING - AN ETHICAL SURVEY

Virginia Horniak
student at the Department of
Computer Science and Engineering,
Mälardalen University
PO Box 883
721 23 Västerås, Sweden
vhk99001@student.mdh.se

ABSTRACT
Privacy of computing is an important ethical topic in today's society. Computer technology, with its many aspects, like for instance communication passed on computer networks, has raised new ethical problems concerning privacy, since a large amount of personal information passes computer networks each day.

Communicating parties have different levels of privacy, depending on what service they want to achieve, but a communicating party has its own rights and duties towards the other parties. Unfortunately a person using the Internet cannot always protect his/her information from everyone, which results in an invasion of the person's privacy.

Privacy has been relevant throughout the history in different ways, but although it is known that privacy is something people are in need of, there are different definitions of what privacy is and how it should be regulated.

This paper gives an overview of computer ethics and specifically the ethical issue of privacy.

1. INTRODUCTION
Privacy is one of the ethical issues which are necessary to take into account when using computer technology. The history of privacy stretches far back and the approach to privacy has changed throughout the times. Today, computers have arised new privacy problems because of the technology, which among others facilitates communication.

Not everyone finds privacy to be something as important as others find it to be and privacy is not perceived in the same way all over the world. Countries have both different opinions of what privacy is and different legislation concerning privacy, which might cause a problem. Since computer networks, like the Internet, are vast, many countries are affected. When all the countries, with their different views of privacy meet, a clash might emerge. Which country's view is the correct one? This type of question might turn up when reading this paper which tries to give an insight in the different aspects of privacy problems.

In the second chapter there will be an introduction to computer ethics and discusses whether computer ethics should be a unique field of ethics or whether all the questions concerning computer ethics already are answered in other fields of ethics. As a result of the discussion a definition of computer ethics is given and the binding between computer technology and computer ethics is described.

The third chapter describes different ethical topics a professional within computer ethics encounters. To give professionals ethical guidelines and to answer some of their questions, professional organizations have formulated their own specific codes of ethics.

In the fourth chapter an introduction to privacy is given. Some confusion might occur between security and privacy, which is given an explanation. When discussing privacy, philosophers have different views a few examples of these are given.

The fifth and the sixth chapter describe the history of privacy and the changes which have occurred. Not all countries in the world feel the same when it comes to privacy and some of the differences are described in the fifth chapter.

The seventh chapter questions the necessity of privacy. Both costs and values with their disadvantages and advantages come with privacy and these are described in this chapter.

An individual communicating with the rest of the society might choose different levels of privacy with different groups of people. The characteristics of these groups and the rights and duties of the groups toward each other are described in the eighth chapter.

2. COMPUTER ETHICS
Computer ethics is an area within applied ethics, where questions related to computers raise new types of moral dilemmas, to which it is necessary to apply the best moral judgement [1]. The society has historically evolved from an agricultural society through an industrial, to present day information society where computers have changed the way people live and make decisions. This type of society has opened doors for new ethical questions never faced by humans before and these questions increase in number along with the development of the technology.

2.1 Uniqueness of Computer Ethics
Whether computer ethics is an independent field of applied ethics or if it can be included in an already existing field, has been argued by traditional ethicists and advocates of the uniqueness thesis in what is called the "computer ethics is unique"-debate, or simply the CEIU-debate [2]. The traditional ethicists do not think there is anything unique about the moral problems which are considered by computer ethicists. These new moral issues like privacy, free speech, intellectual property etc, which are associated with computing, can according to the traditional ethicists be analyzed by using the traditional ethical theories and categories of morality. The mentioned moral issues have not become a new problem in the society because of the introduction of computing; they have all been discussed before and are not necessarily bound to computer ethics. The topics have already been debated and considered by applied ethicists, who have found

a deeper connection to fundamental moral categories, like for example justice, rights, value of life etc.

The opposite side in the debate consists of the advocates of the uniqueness, who find that computer ethics is unique since moral issues concerning computer ethics did not exist before the introduction of computing. Computers have given rise to several new ethical issues like cyber-crimes which were not possible to commit before the introduction of computing. These new ethical issues can be seen as a moral concern, forming a controversial topic like privacy or intellectual property etc.

Some advocates mean that the moral problems associated with computer ethics cannot be analyzed with the traditional morality or the standard ethical theories like, because these ethical theories are not able to handle ethical problems within computing; instead there is a need of a new computer ethical theory or a new framework of morality.

Several known philosophers and computer ethicists, like Deborah Johnson, James Moor and Luciano Floridi and Jeff Sander, have expressed their specific approaches to issues in the CEIU-debate [2]. These views give an understanding of why computer ethics should be a field of applied ethics with a philosophical analysis, but none of the philosophers mentioned have chosen to take side in the CEIU-debate.

2.2 Definition of Computer Ethics

It is important to understand how computer ethics is defined by some known computer ethicists and philosophers, and what kinds of issues are treated when computer ethics is discussed.

Authors writing about this specific ethical issue have different opinions of what computer ethics includes and what it treats. In an early paper about computer ethics, "What is Computer Ethics?", written by the philosopher James Moor [3], a definition of computer ethics is given. This definition is broad and independent of philosophical theories.

Moor defines computer ethics as "the analysis of the nature and social impact of computer technology and the corresponding formulation and justification of policies for the ethical use of such technology". He also states in his paper that computer ethics has no fixed set of ethical rules but instead it considers the relationships between facts, policies and values in a constantly changing computer technology.

According to Moor the introduction of computers and the use of information technology has created "conceptual muddles" and a need of new policies because of the existing "policy vacuums", meaning that there is no fixed set of rules and there are no policies for conduct in certain new situations. The central task of computer ethics is to fill the policy vacuums by formulating guidelines, which are supposed to lead the actions.

In another paper written by Moor, "Reason Relativity and Responsibility in Computer Ethics", the term "logically malleable" is used about computers, which means that computers can be used in many logically different activities [4]. Another term used is "informational enrichment", meaning that computerized settings and activities are constantly developing and becoming informationalized. The fact that computers are logically malleable and that computerized situations become informationally enriched, means that they will generate many new policy vacuums and conceptual muddles or confusions in the future. This also means that the development of computer ethics will never be brought to an end; instead computer ethics is an ongoing process.

2.3 Routine Ethics and Cultural Relativism

Moor discusses how computer ethics should comprise both reason and relativity, since he considers that none of the two popular views called "Routine Ethics" and "Cultural Relativism" is adequate for computer ethics [4]. The view called "Routine Ethics" means that computer ethics is considered as any other ethical area, with no dissimilarities, while in "Culture Relativism" the laws and customs decide what is right and wrong within the field of computer ethics. According to Moor both these propositions are incorrect, because computer ethics needs a discussion and should not be dismissed only by categorizing it into one of these two views.

Instead, computer ethics consists of two parts; the first one is the analysis of situations where computer technology has an impact. The analysis helps to obtain a clear conception of the situation in which policies have to be formulated.

The second part of computer ethics is, according to Moor, the policy-making for using computer technology ethically. The policy-making means that it is necessary to interpret the situation, and to be followed by the evaluation of the policy depending on the society's values system.

2.4 Division of Issues within Computer Ethics

Ethical issues within computer ethics can, according to the philosopher Deborah G. Johnson, be divided into three groups [5].

The first group concerns the ethical issues according to the type of technology they refer to. There has been a large increase of the usage of computers and databases, which are used for record-keeping and the creation, maintenance and manipulation of great amounts of personal information. The development of computer software has raised ethical issues, regarding property rights and the accountability and reliability of programs. Each development in the history of computers, for instance the Internet, has raised new moral concerns.

The second group consists of the ethical issues according to the sector in which they occur. When discussing privacy in general it is for example important not to forget about the different connections, which are protected by privacy, for example the privacy protection of medical records.

The third and last group concerns the ethical issues, according to ethical concepts or theories, where the different ethical issues can be seen from different philosophical points of view, such as privacy, virtue, duty etc. Although there are several alignments in ethics, for example utilitarianism, social contract theory and deontological theory, the theories have a common goal and that is to enhance the man dignity, happiness and the well-being, and to prevent harm. With the help of ethical principles people can achieve this goal for themselves and for other people in different situations.

2.5 Computer Technology and Ethics

To be able to understand the connection between computer technology and ethics, it is essential to recognize the connection between the technology and a human being [5]. It should be pointed out that technology does not yet do anything independently of a human being, but there are situations when the control of a human is weakened when it comes to technology. Especially in those situations it is important to remember the responsibility human beings have for technology, when developing new products. It is essential to keep all the different aspects of a product in mind, especially those affecting the well-being of other people, like safety, reliability, privacy etc.

The codes of ethics, which will be described later, may help computer professionals by guidance.

The understanding of the connection between computer technology and ethics can be divided into two steps. The first step is the acknowledgement of the connection between the computer and the human being.

The second step is to connect a human's actions with ethics. This can be more complicated than it sounds, since computers change the behaviour of human beings by giving the possibility of turning simple movements into very powerful actions. Nowadays, people can with some help of the computer, do things they could not do earlier, and actions can be done in different ways.

3. PROFESSIONAL ETHICS

Another definition of computer ethics is a strict category of professional ethics [6]. Just as occupational groups within medicine and law, computer professionals have their own professional ethics which concern both consumers and the professionals themselves.

Ethical issues, for instance the responsibility towards the customer and the relationship between a computer professional and the customer, client, co-worker, employer and all the other affected people, that a computer professional or a person who professionally uses computers is faced with professionally, are all a part of computer ethics. A professional faces several situations on a daily basis where the ethical values have to be considered, like for example decision-making, where the consequences are significant for both the person who makes the decision and the people who are affected by it.

The ethical decisions made in a professional context may be based on different ethical principles and theories; the virtue ethics, duty ethics (deontological), utilitarian ethics or social contract theory, all of which have a long tradition.

The theory of ethics is based on the assumption that people use rational judgement, making their own choices. This is not always the fact since decision-making is often affected by emotions and other similar circumstances. These emotions are not automatically negative; utilitarian ethics, for example, maximizes the utility/happiness without assuming that people act rationally, but instead strive towards maximum happiness/well-being.

Professional ethics differs from general ethics in the way that a professional has specific knowledge, for example computer science, and the customer (society) does in general not have that knowledge. This leads to a trust where the society must be able to rely on the expertise and the honesty of the professional.

Computer professionals can cause damage through their work, by acting unethically, and customers can have difficulties in protecting themselves from that kind of damage. Therefore it is important that computer professionals have responsibilities to the general public who uses their products or otherwise are affected by the result of their work.

3.1 Social Contract Theory

In order to understand and justify the social responsibilities of a computer professional, computer ethics has automatically been drawn to the philosophical concept of social contract theory [5]. The social contract theory says that since there is a connection between an occupational role, for instance computer professionals who are responsible for the effects of their work, and social responsibilities, a social contract between the professionals and the society in which they work exists. In exchange for accepting certain social responsibilities stated by the society, the professionals obtain numerous privileges from the society. These privileges could for instance be social status, the possibility of forming professional organizations and using educational institutions for education.

The specific ethical responsibilities and requirements concerning computer science have been developed by professional groups working within computer science [6]. The ethical principles of profession, or codes of ethics as they are called in ethics, are in general formulated to be followed by professionals who interact with other people and whose professional activity can affect them.

3.2 Professional Codes of Ethics within Computing

As within many other professions, professional organizations dealing with computer science and engineering have developed a common statement of ethical values, called codes of ethics, which define the specific responsibilities and rights of their profession.

More general codes of ethics treating questions in computer science, are the ACM Code of Ethics [7], developed by ACM (Association for Computing Machinery) and IEEE (Institute of Electrical and Electronics Engineers) Computer Society, and the British Computer Society Codes of Conduct and Practice [8]. "The Software Engineering Code of Ethics and Professional Practice", also developed by ACM and IEEE, are on the other hand more particular, concentrating on the development of software [9].

Certain Swedish trade unions have compiled an own collection of codes of ethics within computer science, called "Ethical rules for computer people" [10]. These codes of ethics are developed by Swedish unions to support their members in situations where conflicts concerning the conscience can arise. The purpose is to increase the member's consciousness dealing with the consequences of their work and to strengthen the confidence in computer technology.

These specific codes of ethics, underline the importance of honesty and fairness in the profession, the maintenance of a professional's competence, the professional's responsibility to respect confidentiality when dealing with computers and the awareness of the legislation that is relevant to the profession. The respect of property rights and the safety of general public are other important topics emphasized in the codes of ethics mentioned above.

The respect of privacy is in the ACM Code of Ethics described as "the responsibility of professionals to maintain the privacy and integrity of data describing individuals" [7].

These codes of ethics are an organization's view of what is crucial considering professional ethics of their members, although the general principles and ideas are the same in all the specified codes of ethics. The members of an organization are obliged to follow specific codes of ethics, while others who are not members might consult the codes voluntarily.

4. PRIVACY

Privacy is next to safety and reliability probably one of the issues worrying people using computers the most. Computer technology allows nowadays a fast and complete search of personal information in computer networks and also surveillance of persons without their knowledge, which to a certain degree was not possible before. New possible threats of privacy have been raised and old threats have become current [1].

4.1 Privacy and Security – Definitions and Differences

Privacy is a relatively new and wide ethics field, including concepts like public/private, confidential/transparent as well as freedoms, rights, security and safety.

The two concepts privacy and security often overlap since they are closely related; however there are some quite important differences between these two issues, for instance conflicts between group security and individual privacy, which are common [11].

The privacy on the Internet concerns the fact that users of it are often worried about loosing their personal information to companies that later on may abuse the information. When discussing personal information, privacy is generally defined as the right of people not to reveal information about themselves, and the right to keep personal information from being misused. The security on the Internet and in computer networks, concerns the communication which can be accessed and manipulated by unauthorized intruders, who have no right to the information passed during the communication.

The main difference between the two concepts security and privacy in computer systems is that the information is secure if the owner has control of it. On the other hand, the information is private if the subject of the information has control of it. Privacy does not only concern information and the risk of loosing control of it; it also concerns private space and private objects which are an important aspect of the personal integrity [12].

Security may be confused with privacy because of the fact that secure, or confidential information, is not open for unauthorized parties, while private information is not revealed without permission.

Anonymity is a term which combines security and privacy by guaranteeing privacy, since anonymous information has no subject. This requires security so that the anonymous information proceeds being anonymous.

The four most important goals for the security of information are:

- integrity, which means that information cannot be changed during transmission
- authentication, which occurs when an identity is established between two users
- confidentiality, which means that the information stays confidential during transmission and
- non repudiation, meaning that it is important to be able to prove that a message has been sent.

4.2 Control Theory and Restricted Access Theory

Within information privacy there are two main theories: the control theory and the restricted access theory [13].

The control theory of privacy was advocated by the philosopher Charles Fried, who meant that people have privacy if and only if they have control of their own information. In other words, privacy is not the lack of personal information about someone in other's minds.

Fried found privacy to be essential to the fundamental good of a friendship between two persons. If there is no privacy, there can not be any confidence in the friendship.

Ruth Gavison's opposing definition of privacy is an example of restricted access theory. The restricted access theory gives a deeper understanding of what privacy is, by defining privacy as the limitation of others' access to an individual, based on secrecy, anonymity and solitude. The term secrecy signifies the limited distribution of personal information, while the term anonymity refers to the protection of unwanted attention and solitude is the decreased physical nearness to others. These three elements should, according to Gavison, be applied by all the participants in the civil society.

James Moor is not of the same opinion as Charles Fried. He means that it is impossible to be able to control own personal information in a computerized society [14]. Instead, it is essential to be sure that only the right people at the right time have access to personal information.

This theory, due to Moor, is his own version of the control and restricted access theory, which he calls the "control/restricted access" theory of privacy. His theory brings control and restricted access theory together, by having all the advantages of the control theory in the restricted access theory. This means that people should have as much control over their data as possible, while it is necessary to keep a focus on the topics indispensable, to be able to develop privacy-protecting policies.

Privacy-policies should be adjusted to the context at the specific time and situation. According to Moor, privacy should not be seen as something definite, "either I know or everybody knows" [14]. Instead, it is a complex composition of situations where personal information sometimes is accessible by some people and sometimes it is not.

4.3 Different Opinions on Privacy

The discussion concerning privacy and what value it has within the professional group described earlier, is a constant issue in focus among ethicists and philosophers [1]. Some definitions have been made and some of the most important ones will be mentioned in this paper.

In the article "Harvard Law Review", written in 1890 by Samuel Warren and Louis Brandeis, the two jurists defined privacy from the legal point of view as "the right to be let alone" [15]. They found that the alterations in the technology and in law were changing the nature of privacy. According to Warren and Brandeis there are two struggling forces in the society; there is the individual who wants to have the right of self-government, or autonomy, and there is the society who wants its members to participate and be active in the society. This was a first basic definition, but there is more to the term privacy than simply the struggle between being left alone and being forced to participate in a society.

Not everyone agrees with Warren and Brandeis analysis of privacy [16]. Privacy is by others seen as a primary good, more similar to a personal dignity, instead of merely instrumental good or a fundamental liberty.

The important difference is how personal information is revealed instead of what private information is revealed.

A way of seeing at the evolution of the right to privacy is that it has followed the development of humanist traditions [17]. Everyone is said to have an intrinsic value, meaning that each human being is valuable in and of him/herself, from which human rights evolve. This development has created a tapestry of privacy where the individual and his/her society has woven together social and psychological qualities with technology and policy. This tapestry of privacy does not give one single meaning to privacy in a society but rather a weave, which covers certain privacy issues in the society.

The view of privacy as a tapestry relates to Warren and Brandeis analysis, by saying that in any case individuals are dependent of society, whether they want to or not. Society has demands on its members. In return the individual has rights and privileges of membership.

In the traditional society, the relationships were often established between the individual and the whole community, without dividing the community into several smaller parts. The technological and social development of the society has led to a renegotiation of the need of individual privacy and the participation in the social community.

Today society looks different and the individual does not only have a single relationship with society, but many relationships with different groups, for example the government, the employer, the insurance company etc. The tools for securing privacy have already been developed. The challenge today is to find a balance between the relationship with the society and the need of privacy. The essential question is not whether an individual has the right to privacy or not, but rather where the line of privacy should be drawn and which tools should be used to realize the privacy.

4.4 Philosophers' Opinions of Privacy

During the Enlightenment in the 18th century, philosophers like Immanuel Kant emphasized that human beings have an ability to communicate and to create [1]. The conclusion was made that if someone's privacy was invaded, the ability to communicate and to create was reduced. That meant that the humanity was automatically decreased. Kant was one of several philosophers during that time period, saying that a person should not give up his/her natural rights when joining a community.

According to Kant, people in powerful positions, should "act so as to treat human beings always as ends and never merely as means", meaning that every person has individual purposes like goals, desires or choices and if a person is treated as means, (s)he looses the ability to decide for him/herself and becomes enslaved. When a person is treated as an end his/her goals, desires and choices are fulfilled, meaning that if someone is treating a person as an end, (s)he is striving towards the person's purposes.

Kant also found the individual to be rational and autonomous; therefore the choices an individual is making have to be private and personal.

John Locke went a step further, saying in the Second Treatise of Government that if someone threatens an individual's property, which is private, (s)he potentially threatens the individual's life. It gives the individual enough reason to even kill the person who is threatening the property of an individual, when protecting his/her life.

Locke compared the defence of private property and the right of an individual to own his/her own body. When an individual's private property is threatened, parallels can be drawn to this individual's non-physical privacy, for instance the privacy of communication or the right to free speech. Locke's statement comprises many of the ethical questions about right to privacy.

Locke's opinion, giving an individual the right to kill another person who threatens his/her property or non-physical privacy, is a strong statement which opposes the universal judgements of behaviour [18]. These universal ethical principles, which are caring for children, trust and prohibition against murder, can be seen as basic principles which apply to all cultures. When a known philosopher like, John Locke, opposes the three core values the conviction he has, that privacy is a fundamental right, strengthens the position of the advocates of privacy and their belief in importance of privacy.

A true Utilitarian might say that an invasion of privacy causes more harm than good. When private information about a person has been spread, either it is false or true, there has been an invasion of the person's privacy [1]. The aggression can be discussed depending on the context; the wiretapping of a terrorist or a criminal is ethically defensible because of the need of security in a society.

James Moor comes to the conclusion that privacy is a necessary condition for an intrinsic good, since it is possible to assume that autonomy is intrinsically valuable and privacy is a necessary condition for autonomy [14]. Moor does not find privacy to be only an instrumental value, which means good leading to something else that is good, as many other philosophers find it to be; he claims that privacy is intrinsically valuable. An intrinsic value is a value which is good in itself.

Moor states further the importance of privacy by saying that human beings have "core values", which are fundamental to the human evolution. These core values are for example life, happiness, freedom, knowledge, ability, resources and security, and each core value is emphasized by some people more than others. Although he does not find privacy being a core value, he thinks privacy is an expression of security, which is a core value. Without privacy it is difficult to feel secure and evolve as a human being.

5. PRIVACY IN OTHER PARTS OF THE WORLD

In some parts of the world, for example in Sweden and in Japan, there is no single word for "privacy" as in the English language. In Sweden for example the word "privacy" has been replaced by the words "personal integrity" [19], while in Japan one of the many adjacent words to "privacy" is the word "secret" since there is no specific word for "privacy" in the traditional Japanese language [20]. The topic discussed by the authors of the paper "The internet and Japanese conception of privacy" concerns the fact that Japan has a different point of view regarding privacy because of its cultural, linguistic and historic development. Dismissing privacy as a concept in Japan, because of the lack of word is to simplify the problem, because the concept exists and therefore it is possible to discuss the subject.

Considering the population density and the living conditions in Japan, as a consequence of a compact style of living, there is a difficulty in preserving privacy in an everyday-life. Remember that homes in Japan are separated into rooms by traditional lattice-work doors. Even when those doors are closed, they do not provide any considerable physical barrier. There is an unwritten law in Japan saying that even if people gain information in a situation where they overhear information due to the standard of living, they should act as if they have not gained this information.

In a society, which has a lack of some privacy, there is something called normative privacy, protecting people from intrusion. For example, even if a person is walking in a public street, it is forbidden to look under this person's clothing. This normative privacy is apparent in Japan, although the details of privacy are different from those in Western cultures.

Belonging to a group and working together responsibly characterizes the Japanese culture, while the individual role, even if belonging to a group, is for example more emphasized in the US. The view of privacy differs between a Japanese individual and an individual in the US, but according to the authors of the paper "The internet and Japanese conception of privacy", there is a basic and a common understanding of privacy in any developed culture, which is called the minimal conception of privacy. The culturally developed privacy in individual countries is called the rich conception of privacy, and that is what mainly differs between the Western countries and Japan.

5.1 The Extension Problem and the Coordination Problem

Since the definition of privacy varies between countries, it might raise two different problems in the policy-making: the extension problem and the coordination problem.

If a society's norms have to expand to be able to handle new situations because of new technologies like the Internet, then the actual society has an extension problem.

The society has a coordination problem when its norms collide with the norms of another society. A society might for example treat personal information on the Internet in one specific way, while another society treats it differently. The coordination problem that comes up, is which society's norms should be applied. In the paper "The internet and Japanese conception of privacy", a new idea comes up, and that is to develop a common view of the rich conception of the Internet privacy, where each user's personal information would be protected.

5.2 Privacy Today

The question at issue concerning privacy has changed throughout the times. As mentioned earlier, the problem is not if the individual has right to privacy, but where the line of privacy should be drawn, for example if the authorities should have access to encrypted information [17]. Due to the attack of the September 11, 2001, the question has once again changed. Today the question is instead how much of the information the authorities should have, knowing that the terrorists used computer networks to communicate and to plan the attack [21]. If their communication would have been revealed, the attack could have been prevented, but since this was not the case, the debate concerning privacy has changed direction and the opinion against wiretapping softened directly after the attack.

6. HISTORY OF PRIVACY

Apparently, the history of privacy stretches far back in time and throughout time privacy has been an important right in societies. Although there are opinions saying that privacy was not relevant in non-democratic and non-technical societies in other sense than having a physical privacy by being able to withdraw from other people [22].

Societies in the pre-state or tribal stage of development, for example Eskimo and American Indian societies, were often politically unstructured and there was no social boundary between private and public [22]. Often everyone knew everyone else and no one was anonymous. Ancient states, for example Greece and China, were well-developed societies with elements of privacy, but otherwise privacy concerned first highly organized and large societies where that kind of relationship between the individual and the rest of the society could not be directly negotiated, but instead must have been institutionalized.

The installement of the centralized public church, and the imperial and royal authorities did not introduce privacy as a social issue; instead they were powerful and provided stability [22]. The society was at that time distinctly stratified, where the institutions in power, as the clergy and the royalties, were in the top layer, and the peasants were below. The lower layer of the society did in reality not have many rights, and especially no right to privacy. Even the upper class and the royals lacked of privacy; royals belonged to the public and each event in their lives was observed by others, either by the court surrounding them, or by the general public.

6.1 Change of View Regarding Privacy

Things changed and, as already mentioned, the Enlightenment which occurred in the late 17^{th} century had a large influence on people's way of thinking of their right to privacy [22]. The Enlightenment changed the individual's view of the rights, and his/her position in the society. At that time, people understood that their experiences were important and that the result of their experiences was new knowledge. Earlier the upper class controlled and treated the lower class as a mass of people instead of treating them as individuals. Immanuel Kant and other philosophers came during the Enlightenment to the conclusion, that the mass should be treated as individuals and that each and every one of the individuals have the right to determine what the truth is, instead of being forced by the upper class to believe in its already defined truth.

Another impact, which changed the individual's approach to the right to privacy, was made by the growth of the mercantilist class and the bourgeoisie in the 18th century. These classes grew stronger and took control of the economy in the society, which earlier was in the hands of the clergy or the kingdom. By controlling the economy the political power came; this included the privileges and rights. It was important for the new classes to protect their new assets from the public state. Therefore they got political and established a legal protection of their property, which led to the first protection of privacy. The mercantilist class and the bourgeoisie got the political and legal rights to protect and control their private assets, without any interference of the state.

6.2 The Impact of Computer Technology

The definition of the right to privacy has changed together with the developing computer technology, the storage of information in databases and the passing of information in computer networks [22]. The concern about government, holding information about citizens as a secret, led to legislation. Many people see information as something that should be accessible by the public, and not kept in secret by governments. The legislation gave the public right to access some insensitive information. Soon there was a concern about misuse of personal information which the public had access to. This led to additional legislation which specified how personal information should be gathered and used by the government or within a company. This also provided the individual a possibility to confirm that the information about him/her was accurate.

The evolvement of the right to privacy has been dependent of a country's history, which has led to different interpretations of boundaries between personal and state rights. Attitudes towards the right to privacy, freedom of information and the legislation varies therefore between countries. It is not a fact that the right to privacy should be a self-evident right.

7. WHY PRIVACY?

Before the development of telecommunication and computer technology took place, information was often spread by direct verbal communication, including gossip. When personal information is taken out of its context, there could be a risk of misjudgement of a person. When hearing information about someone else, impressions of this person are often formed, and they can either be true or false.

7.1 Is Privacy Really Necessary?

Privacy protects people from being judged by others due to possible false information spread by others [23]. According to some, for instance the philosopher Charles Fried, true knowledge about an individual can only be achieved by some close persons related to this individual, where the individual has the right to choose the degree of intimacy in a relation with other people. To be able to grow a close relationship, there is a need of privacy and this privacy excludes the surroundings which have formed an opinion about the individual based on sometimes false information. The individual must however interpret which characteristics (s)he wants to be defined by. That is enabled through his/her rights privacy in the sense of the control of ones own personal information.

Privacy is not always seen as something necessary. The social value of privacy is questioned [23]. Some people say there is a risk that privacy makes people disguise true information about themselves to gain advantages in the social or economic life. Another opinion is that having a private life, in addition to the public life, is a social fraud and leads to deception and hypocrisy. Saying this is, according to the defenders of privacy, confusing privacy and secrecy which is only a small part of privacy.

7.2 Masks of Privacy

Concerning the differences between the public and the private lives leading to a social fraud, there is a parallel drawn to persons wearing different "masks" depending on which situation they are in. If people are in a public situation, they wear one type of "mask", and if it is a private situation, another "mask" is worn. The defenders of privacy mean that if this "mask", containing people's different characteristics in the specific situation, would be removed the person behind the "mask" would not be his/her true self but instead a defenceless person, not prepared for the present situation (s)he is in. As an example the behaviour of an influential executive who plays two different roles, depending on whether (s)he is at the office in an official context or if (s)he is at home playing with the children. In general, people play different roles on different occasions and the "masks" they are wearing are only an expression of the different sorts of relations they are having with different people.

7.3 Social and Personal Costs and Values

Opinions mean that privacy has political, social and personal costs and values attached, if it is defined as the capability to protect oneself from judgements based on false information.

The political value involves the fact that there is no need of revealing one's rank or family background to interact with others in a democracy. Because of privacy, there is a possibility for the citizens, who might disagree on a topic, to interact with each other without having to reveal their identity.

The social cost could for example be surveillance, which is common in both public and private workplaces. There are figures saying that one-third of the online working workforce in the US, is under surveillance [21]. Surveys show that the employees are often more depressed, tensed and anxious, knowing they are monitored, than those who are not under or unaware of surveillance [23]. Philosophers mean that it is obvious that a surveilled person behaves differently than a person who is not monitored, because the monitored person is aware that his/her opinions and actions are watched by a third party, the employer who performs the surveillance. There is a risk that the communication, for example emailing between employees, becomes less efficient because of the increased formality in the communication.

The personal cost of privacy is the commitment to privacy [23]. Charles Fried says that human feelings, for instance respect, love and trust are unimaginable without privacy, meaning that intimacy and privacy are essential parts in relationships, either friendly or romantic. The degree of intimacy depends on the amount of selective personal information which has been revealed.

Privacy is described as a moral opacity, which creates a balance between a transparent society and a society where its citizens are totally isolated from each other.

8. PRIVACY AND RELATIONSHIPS

Privacy might be described as focused on the relationships of four groups [17].

The first group consists of an individual, for example a student, who has the right to privacy, both to physical privacy and to the protection of personal information [17]. Individuals share their information in return for relationships or services from the second group, for example a network administrator.

The third group does not directly receive the information shared between the student and the network administrator. This group has access to the information about the student as a consequence of their professional role. The information should not be used, since they are involved in activities which are irrelevant to the student, who is not even aware of the fact that the third group might have information about them.

The fourth group consists of all the rest of society also known as the public, which has access to some of the public personal information about a student, which is “out in the open”.

8.1 Rights and Duties

Each one of these four groups has its own rights and duties towards the other groups [17]. Individuals have the right to privacy, but it is never an absolute right. Individuals have, for example, not only the right to stop, restrict and control their public, personal information, but they have the responsibility to be active while protecting their rights. There are forces, for instance authorities, who can interfere with the individual’s will of privacy. One could think that a police-search of the premises, or a personal search, is something privacy-intrusive but still the police have certain rights to perform their searches and the individual’s right to privacy can be disregarded under certain (legal) conditions.

8.2 Levels of Privacy

Depending on the interaction between different groups, individuals can invoke different levels of privacy [22]. The advantages have to be compared with the risks dealing with the release of information. Finally, individuals make choices dealing with the sort of information which will be exchanged. This choice is based on the balance between the advantages of the release of information, and the risks of inappropriate use of it.

Individuals should achieve information about the second group before creating a relationship with it. Individuals have to be aware of what sort of information they have to provide, and how this information later on will be used. This type of relationship is called a negotiated relationship.

The second group, the network administrator, is in a trustful relationship with the student and therefore should the personal information be secured by the network administrator in such way that the third group does not access it [17]. The student must confide in the network administrator and often there is a moral problem when identifying the rights of the network administrator, i.e. who has the right to the personal information and who has not.

The third group should not try to access the individual’s personal information, and respect the individual’s right to privacy. There are differences, as described above, between cultures considering the privacy definitions. In cultures, for example Japan, the view of privacy is different compared with Western countries. In Japan the convention says that even if a third group would gain information about the first group, the student, in a certain situation where the information was not supposed to leak out, the third group should act as if the information was not available to them [20]. An example is the network administrator who has access to private information about the students, but (s)he is supposed to act as if (s)he did not have access to it.

Finally the fourth group should only have access to the information they are entitled to, being able to perform their social functions [17].

9. CONCLUSION

This paper has given the reader an introduction to computer ethics and an overview of privacy.

The different opinions of what privacy is and if it really is necessary in our society have been analyzed. The opinions are many all over the world; each country has its own legislation concerning privacy, making it difficult to combine it on the Internet where communication from different parts of the world meets.

The ethical theorists have of course their own views of privacy. Some views are more direct than others, but most of the philosophers agree on the fact that privacy is something inevitable in our society.

The fact that privacy is necessary has been questioned in this paper. There are both advantages, like social values, and disadvantages, like social and personal costs, which come together with privacy. A person seeking privacy must take these advantages and disadvantages into consideration.

Communicating parties, which are divided into groups with different levels of privacy, also have rights and duties towards each other. These rights and duties should be followed to preserve privacy when communicating.

Only when the rights and duties of these four groups have been settled, a technical problem rises how to design and to implement a system, which preserves the privacy of communicating parties.

10. REFERENCES

[1] Stacey L. Edgar. Morality and Machines, Second Edition. Jones and Bartlett Publishers. ISBN 0-7637-1767-3

[2] Herman T. Tavani. The uniqueness debate in computer ethics: What exactly is at issue, and why does it matter? Ethics and Information Technology 4, 2002

[3] James H. Moor. What is computer ethics? Metaphilosophy 16/4. http://www.ccsr.cse.dmu.ac.uk/staff/Srog/teaching/moor.htm

[4] James H. Moor. Reason Relativity and Responsibility in Computer Ethics. Computer Ethics and Professional Responsibility. Edited by Bynum, Terrell Ward, Rogerson, Simon. Blackwell Publishing. ISBN 1855548453

[5] Deborah G. Johnson. Computer Ethics. The Blackwell Guide to the Philosophy of Computing and Information. Edited by Luciano Floridi. Blackwell Publishing. ISBN 0-631-22919-1

[6] Sara Baase. A Gift of Fire, Second Edition. Prentice Hall. ISBN 0-13-121988-X

[7] ACM Code of Ethics and Professional Conduct. http://www.acm.org/constitution/code.html

[8] British Computer Society Codes of Conduct and Practice. http://www1.bcs.org.uk/link.asp?sectionID=673

[9] ACM Software Engineering Code of Ethics and Professional Practice. http://www.acm.org/serving/se/code.html

[10] Etiska regler för datafolk. Dataforum. http://www.dataforum.org/pdf/etikbok.pdf

[11] Richard A. Spinello, Herman T. Tavani. Readings in CyberEthics. Jones and Bartlett. ISBN 0-7637-1500-X

[12] L. Jean Camp. Web Security and Privacy: An American Perspective. The Information Society 15, 1999

[13] Richard A. Spinello. Cyberethics Morality and Law in Cyberspace, Second Edition. Jones and Bartlett Publishers. ISBN 0-7637-0064-9

[14] James H. Moor. Towards a Theory of Privacy for the Information Age. Computers and Society, Vol. 27, No. 3, 1997

[15] Warren and Brandeis. The Right to Privacy. Harvard Law Review, Vol. 4, No. 5, 1890

[16] Paul. B. Thompson. Privacy, secrecy and security. Ethics and Information Technology 3, 2001

[17] Richard O. Mason. A tapestry of Privacy, A Meta-Discussion. http://cyberethics.cbi.msstate.edu/mason2/

[18] James Fieser. Moral Philosophy through the Ages. Mayfield Publishing Company. ISBN 0-7674-1298-2

[19] Anders R. Olsson. Efter 11 september 2001: - Kan Storebror hejdas? Vinnova & teldok. http://www.vinnova.se/main.aspx?ID=035C493E-67E7-43E4-8FFE-9BF92EB4F786

[20] Masahiko Mizutani, James Dorsey, James H. Moor. The internet and Japanese conception of privacy. Ethics and Information Technology 00, 2004

[21] Stephen Hinde. 2001: A Privacy Odyssey Revisited. Computers and Security, Vol. 21, No. 1, 2002

[22] Ian Graham. Putting Privacy in Context – An overview of the Concept of Privacy and of Current Technologies. http://www.utoronto.ca/ian/privacy/privacy.doc

[23] Jeffrey Rosen. Why Privacy Matters. Wilson Quarterly. Vol 24, Issue 4, 2000