Deutsches
Forschungszentrum
für Künstliche
Intelligenz GmbH

Research Report
RR-04-02

Proceedings of the First Workshop on Philosophy and Informatics (WSPI 2004)

Cologne, Germany

Gregor Büchel, Bertin Klein, and Thomas Roth-Berghofer (Eds.)

March 2004

Deutsches Forschungszentrum für Künstliche Intelligenz GmbH

Postfach 2080
67608 Kaiserslautern, FRG
Tel: +49 (631) 205-3211
Fax: +49 (631) 205-3210
E-Mail: info@dfki.uni-kl.de

Stuhlsatzenhausweg 3
66123 Saarbrücken, FRG
Tel: +49 (631) 302-5252
Fax: +49 (631) 302-5341
E-Mail: info@dfki.de

WWW: http://www.dfki.de

Deutsches Forschungszentrum für Künstliche Intelligenz

DFKI GmbH

German Research Centre for Artificial Intelligence

Founded in 1988, DFKI today is one of the largest non-profit contract research institutes in the field of innovative software technology based on Artificial Intelligence (AI) methods. DFKI is focusing on the complete cycle of innovation — from world-class basic research and technology development through leading-edge demonstrators and prototypes to product functions and commercialisation.

Based in Kaiserslautern and Saarbrücken, the German Research Centre for Artificial Intelligence ranks among the important "Centres of Excellence" world-wide.

An important element of DFKI's mission is to move innovations as quickly as possible from the lab into the marketplace. Only by maintaining research projects at the forefront of science can DFKI have the strength to meet its technology transfer goals.

DFKI has about 165 full-time employees, including 141 research scientists with advanced degrees. There are also around 95 part-time research assistants.

Revenues for DFKI were about 30 million DM in 2000, from government contract work and from commercial clients. The annual increase in contracts from commercial clients was greater than 20% during the last three years.

At DFKI, all work is organised in the form of clearly focused research or development projects with planned deliverables, various milestones, and a duration from several months up to three years.

DFKI benefits from interaction with the faculty of the Universities of Saarbrücken and Kaiserslautern and in turn provides opportunities for research and Ph.D. thesis supervision to students from these universities, which have an outstanding reputation in Computer Science.

The key directors of DFKI are Prof. Wolfgang Wahlster (CEO) and Dr. Walter Olthoff (CFO).

DFKI's seven research departments are directed by internationally recognised research scientists:

- ❑ Image Understanding and Pattern Recognition (Director: Prof. Th. Breuel)
- ❑ Knowledge Management (Director: Prof. A. Dengel)
- ❑ Intelligent Visualisation and Simulation Systems (Director: Prof. H. Hagen)
- ❑ Institute for Information Systems (Director: Prof. A.-W. Scheer)
- ❑ Deduction and Multiagent Systems (Director: Prof. J. Siekmann)
- ❑ Language Technology (Director: Prof. H. Uszkoreit)
- ❑ Intelligent User Interfaces (Director: Prof. W. Wahlster)

In this series, DFKI publishes research reports, technical memos, documents (e.g. workshop proceedings), and final project reports. The aim is to make new results, ideas, and software available as quickly as possible.

Prof. Wolfgang Wahlster

Director

Proceedings of the First Workshop on Philosophy and Informatics WSPI 2004, Cologne, Germany

Gregor Büchel, Bertin Klein, and Thomas Roth-Berghofer (Eds.)

DFKI-RR-04-02

ISSN 0946-008X

Preface

The interest in the workshop "Philosophy and Knowledge Management" at the "Second Conference on Professional Knowledge Management WM 03" in Luzern, Switzerland, and continuing interest thereafter fortified our decision to continue with this work. Already in Luzern, we discussed the idea to found a Special Interest Group (SIG) as part of the German Informatics Society GI (Gesellschaft für Informatik). In autumn 2003, finally, we founded the SIG "Philosophy and Informatics" and, as our first major activity, organized this "First Workshop on Philosophy and Informatics WSPI 2004" in Cologne.

Philosophy has a long tradition in Cologne. Its first university was founded in 1248 by the medieval philosopher Albertus Magnus (ca. 1193–1280). Albertus was the teacher of Thomas of Aquin (1225–1280). Students of the medieval University of Cologne had a trans-disciplinary curriculum, including logic, geometry, astronomy, and dialectics (philosophy). This tradition can be seen as a motivating aspect that led to the trans-disciplinary goals of this workshop.

The work of the Special Interest Group "Philosophy and Informatics" addresses and encourages the discourse on foundations of Artificial Intelligence with the help of Philosophy. In this discourse, interfaces between philosophical points of view and points of view of informatics are discussed. With this workshop, first steps to build a common vocabulary of philosophers and computer scientists have been taken and are documented with the present proceedings. For further information about the goals and the results of this workshop and about the progress of the SIG's work, please, visit the SIG's website at http://www.nt.fh-koeln.de/philosophyandinformatics.

Many persons are needed for the success of a workshop like this. We thank the contributors, the program committee's members, and the supporters at the German Research Center for Artificial Intelligence DFKI, Kaiserslautern, as well as those at the University of Applied Sciences, Cologne.

March 2004

Gregor Büchel, Bertin Klein,
and Thomas Roth-Berghofer
Organizing Committee
WSPI 2004

Organization

The "First Workshop on Philosophy and Informatics WSPI 2004" was organized by the GI Special Interest Group "Philosophy and Informatics" (Arbeitskreis "Philosophie und Informatik" des Fachbereichs "Künstliche Intelligenz").

Organization Committee

Gregor Büchel (University of Applied Sciences Cologne, Germany)
Bertin Klein (DFKI, Germany)
Thomas Roth-Berghofer (DFKI / University of Kaiserslautern, Germany)

Program Commitee

Daniela Bailer-Jones (University of Bonn, Germany)
Thomas Börnchen (Brose Fahrzeugteile GmbH & Co. KG, Germany)
Pierre Grenon (IFOMIS, University of Leipzig, Germany)
Ruth Hagengruber (University of Koblenz and Landau, Germany)
John D. Haynes (University of Central Florida, USA)
Norbert Jastroch (MET Communications GmbH, Bad Homburg, Germany)
Jürgen Müller (University of Cooperative Education, Mannheim, Germany)
Michael Paetau (University of Essen, Germany)
Steve Probert (Cranfield University, Swindon, United Kingdom)
Christophe Roche (University of Savoie, France)
Francis Rousseaux (Institut de Recherche et de Coordination Acoustique Musique (IRCAM), Paris, France)
Steve Russ (University of Warwick, United Kingdom)
Barry Smith (IFOMIS, University of Leipzig, Germany)
Marcus Spies (University of Munich, Germany)
Boris Wyssusek (Technical University Berlin, Germany)

Table of Contents

Full Papers

Statement of Interest

Analyzing Knowledge Management Systems: A Veritistic Approach

Palash Bera[1], Patrick Rysiew[2]

[1]MIS division, Sauder School of Business, 2053 Main Mall, University of British Columbia, Vancouver, V6T 1Z2, Canada
[2]Department of Philosophy, 1866 Main Mall, University of British Columbia, Vancouver, V6T 1Z1, Canada
{palash, rysiew}@interchange.ubc.ca

Abstract. Knowledge management systems (KMS) are increasingly becoming popular and important in managing organizational knowledge. This motivates a closer inspection of the degree of usability of various types of KMS. This paper is an analysis of KMS from a philosophical angle: with the help of veritistic social epistemology we analyze which KMS are likely to be used *more* in comparison to others. Veritistic social epistemology is oriented towards truth determination; it seeks to evaluate actual and prospective multi-person practices in terms of their tendency to produce true beliefs (versus false beliefs or no belief) in their users. We distinguish between KMS that manage structured knowledge and those that manage unstructured knowledge. It is argued that structured knowledge is more credible to the users than unstructured knowledge and that, because of this, KMS that manage structured knowledge bring more veritistic gains than those that manage unstructured knowledge.

1 Introduction

The aim of this paper is to explore the question, "Which knowledge management systems (KMS) bring more veritistic gains to users in comparison to others?" Veritistic social epistemology seeks to evaluate actual and prospective practices in terms of how well they tend to promote the acquisition of true beliefs (versus false beliefs or no belief) on the part of their users. The practice of Knowledge Management (KM) is realized by using IT products or information systems (IS) called KMS. This paper analyzes two types of KMS and finds which type is veritistically superior. The paper is organized as follows. The background of the paper is set out in the next two sections, on knowledge in organizations and veritistic social epistemology, respectively. Veritistic analyses of KMS, with the interpretation of the findings are done next followed by the conclusion.

2 Knowledge in Organizations

A commonly held view among IS researchers is that *data* is raw numbers and facts, *information* is processed data, and *knowledge* is authenticated information [1]. Another common, perhaps complimentary, way of thinking about knowledge in an organizational context, is as "information in action" [cf.2]. Nonaka [2] discusses two types of knowledge in organizations: *tacit* and *explicit*. Tacit knowledge is rooted in action, experience and involvement in specific context. An example of tacit knowledge is the artful oral skills of an experienced sales-person selling a not so useful product. Explicit knowledge is articulated, codified, and communicated in symbolic form and/or natural language. An example is a product manual that accompanies a product, describing what it does and how it works. (The distinction between explicit and tacit knowledge corresponds to the distinction between "propositional" and "procedural" knowledge, respectively.) The tacit form of organizational knowledge is related to the production of knowledge by social means. According to Stahl [3], individuals generate personal beliefs from their own perspectives, but they do so on the basis of socio-cultural knowledge, shared language and external representations. Thus beliefs come to be accepted as knowledge through social interaction, communication, discussion, clarification, and negotiation. So understood, knowledge is considered as the product of social processes.

2.1 Knowledge Management Systems

Organizations are collections of humans, their skills and resources. The resources consist of human skills (e.g. expertise, experience), physical resources (e.g. building, machinery) and vast collections of information and knowledge. "Knowledge" in this context includes both the experience and understanding of the people in the organization and the information artifacts, such as documents and reports, available within the organization and in the world outside [4]. Organizational knowledge is difficult to manage, as the volume of knowledge (both tacit and explicit) increases day by day and tacit knowledge embedded in human minds disappears from the organization when employees leave. In order to capitalize on organizational knowledge, organizations must create an environment where knowledge is captured, shared and transferred effectively and efficiently. KM helps to capture, share, and transfer knowledge and thus manage organizational knowledge. The practice of effective KM typically requires an appropriate combination of organizational, social, and managerial initiatives, along with the deployment of appropriate technology. Technology can help to capture explicit knowledge (e.g. in databases), identify sources of tacit knowledge (e.g. using extranet applications), share tacit and explicit knowledge among users (e.g. using groupware) and transfer knowledge (e.g. preparing best practices documents). A KM system is a specific type of IS or IT product, applied to managing organizational knowledge [1].

2.2 Categories of KMS: Structured and Unstructured Knowledge

Hahn and Subramani [5] propose a framework for classifying KMS, based on where the knowledge (to be managed) resides and the extent to which it is structured. They distinguish between KMS in which knowledge is *structured*, and those in which it is *unstructured*. "Structured knowledge" either has an inherent structure (as does, e.g., an electronic database) or the structure is imposed upon it (as, for example, when documents are tagged with keywords) [5]. Structured knowledge is generally stored in the organization such as in the corporate websites. Unstructured knowledge has no prior structure and is generally dynamic in nature for example information in an electronic discussion forum. It is difficult to impose structure on unstructured knowledge. For example, tagging text obtained from an electronic discussion forum is difficult as the content is highly context dependent. Organizational knowledge that is structured can essentially be coded and thus classified as explicit knowledge. Tacit knowledge, which is generally unstructured (such as expertise or experience), is created over a period of time in organizations.

KMS have been developed that manage both structured and unstructured knowledge. Some KMS use classification mechanisms, tags or meta tags to structure knowledge and then manage them. An example of a KM system managing unstructured knowledge is the collaborative filtering systems. These systems predict browse and search behaviors by analyzing past behavior of other users when they performed a similar activity [5].

3 Veritistic Social Epistemology

Epistemology is the branch of philosophy that deals with the study of knowledge. Traditional or classical epistemology is concerned with the pursuit of truth, typically on the part of individuals considered in isolation from other agents or any broader social setting. Social epistemology focuses on the social dimensions of knowledge and/or knowledge-production; it acknowledges, where traditional epistemology largely ignores, the important role that social factors play in the knowledge-forming process. This is not to say, however, that social epistemology must jettison the traditional epistemological concern with truth (true belief): *veritistic social epistemology* (VSE). VSE is concerned with the role of social factors in 'the production of knowledge, where knowledge is understood in the ''weak'' sense of *true belief*' [6] (p. 5). VSE is intended to be evaluative or normative rather than purely descriptive or explanatory: the task of the theorist is to identify and evaluate actual and potential social processes/activities/institutions in terms of their tendency to promote the acquisition of true belief (versus false belief or no belief) in their users [7].

The motivation for adopting a veritistic approach is straightforward: both for practical reasons and because they are spontaneously curious, humans across cultures and throughout history commonly seek the truth. Moreover, epistemic notions such as *knowledge* are properly conceived in terms of truth: when we discover that a belief is false, we cease to consider it to be a candidate for knowledge -- something one might "know" -- in any non-figurative sense. From this perspective, when "knowledge" is used to refer to *what is accepted* within an organization, this must be taken with a grain of salt: if what is accepted turns out not to be true, then it is not knowledge, properly so-called, even though users within the organization might continue to consider it to be such. Hence, understanding and evaluating the role of inter-personal, *social* factors in knowledge acquisition/production requires that we look at their tendency to promote true belief. In a word, institutions and practices that foster *true* belief are epistemically good and should be promoted; institutions and practices that result in false belief (error) or the absence of true belief (ignorance) are epistemically bad, and should be avoided or corrected. (This is assuming that there are not overriding, non-epistemic reasons which speak against a given veritistically good practice or institution, or in favor of some practice/institution which, while less good epistemically, is judged to be better overall.)

The main question for VSE is thus, "which practices have a comparatively favorable impact on knowledge as contrasted with error and ignorance?" [6] (p.5). Once again, the rationale for taking up a veritistic perspective is that in everyday life a certain value is placed on having true beliefs rather than false beliefs or no opinion. This type of value is *veritistic value* (or "V-value"). In order to understand the concept of V-value we cite an example from Goldman [6]. Suppose that a person S has an interest in a yes/no question as: "Is it the case that P has occurred?" V-values can be assigned for three possible states. If S believes that the proposition is true then the V-value is 1.0, if he rejects the true proposition P then the V-value is 0, and if he withholds the judgment then the V-value is 0.5. The first case constitutes *knowledge*, the second *error* and the third *ignorance* respectively. Veritistic analysis focuses on change of V-value over time. Over a period of time if a person changes his state of belief from no-opinion to rejecting P, then the V-value either improves or worsens depending on whether P is true or false.

A high V-value of beliefs indicates an increase in the level of knowledge of the user. If the V-value increases of an entire community then the knowledge of the entire community increases. *The increase in V-value can yield veritistic profit to the users whose beliefs have been modified.* If a user moves from false belief to true belief by receiving correct information, and the receiver is able to draw true conclusions from the information, then there is an increase in veritistic profit of the user. Finally, it is important to note that veritistic analysis is always to be assessed relative to the questions of interests. Though they may be true, answers to questions that are of *absolutely no interest* to the user do not qualify as properly having V-value.

Knowledge in organizations is formed, shared and practiced by processes that are essentially social. The practice of managing knowledge in organizations is done by KM, and this practice is realized by KMS. KMS help to bring changes to individual beliefs by allowing individuals to use KMS. Any tool such as a KM system will have an impact on modifying the beliefs of users, resulting in change of V-value. However, we predict that not all KMS will have an equal impact on modifying users' belief. This is because different KMS manage different types of knowledge (structured and unstructured). Users might be influenced by the credibility of these knowledge and modify their beliefs accordingly. In the next section we analyze the credibility of these different types of knowledge, and their consequent impact on change in V-value.

4 Veritistic Analysis of KMS

4.1 Credibility of Knowledge Sources

We make a claim that: "*structured knowledge is more credible to the users than unstructured knowledge*". This claim is based on analyzing a similar proposition of credibility between structured and unstructured information. Information is normally seen as that which has meaning, in that it reduces uncertainty for the seeker [8]. Information can, however, increase uncertainty and create more dissonance [9]. Information can also be viewed as something that describes fact. Hicks et al. [10] mention that individuals exposed to structured information may infer the same knowledge from it and majority of unstructured information is either personal or developed through interaction between two or more individuals. For example, "an engineering drawing" is a structured piece of information containing text, numbers and symbols and drawn for a specific context. This information can also be evaluated for its accuracy (thus measurable). Different engineers will infer similar knowledge by studying the drawing. The primary differences between structured and unstructured information are outlined in Table 1.

Table 1. Information category [10]

Category	Characteristics
Structured	• Textual (e.g. numeric, alphabetic) and pictorial (visual image) • Generally context dependent • Consistent by producing similar knowledge from structured information
Unstructured	• Textual (e.g. personal note), verbal (conversation) and memory • Generally context independent • Inconsistent as individuals infer different knowledge from unstructured information

Trust is an important factor in belief formation. McDowell [11] points out that trust can have a crucial epistemic impact on social epistemology, which assesses the epistemic value of social practices. Foley [12] points out that our most fundamental assumptions from where opinions are "formed" and not "self-generated". They are passed to us as part of our intellectual inheritance. For example, we do not verify the assumptions mentioned in the elementary science text books, as (we think) the assumptions are already verified by some experts and these assumptions often become the basis of our or others beliefs. Thus, formation of our fundamental beliefs depends on the trust we place in their sources. Hardwig [13] (p. 694) mentions that "trust is often epistemologically even more basic than empirical data or logical arguments: the data and the arguments are available only through trust".

Trust and credibility are closely related. Credibility is defined as "believability" and trust is regarded as "a positive belief about the perceived reliability of, dependability of, and confidence in a person, object, or process" and in this sense trustworthiness of information is a synonym for credibility [14]. "Credible" sources are described as "trustworthy" and having "expertise" [15]. McDowell [11] points out that an epistemic effect of social trust is that people can be less willing or likely either to speak or to listen to socially untrustworthy sources of information, and more willing or likely either to speak or to listen to socially trustworthy ones. Self presents a summary of early Socratic and Aristotelian ideas about credibility [15]: "First, sources are credible because their message's rightness is perceived by the audience. Second, sources are credible because they rightly read how to reveal themselves to particular audiences. And, third, sources are perceived to be credible because of audience characteristics." (p.423) In most cases a message's rightness is more easily perceived by the users using structured information. Users are able to rightly deduce the meaning of the structured information more consistently than unstructured information. For example, studying an engineering drawing of an engine (structured information) versus listening to the functioning of an engine (unstructured information), different users would be able to interpret the drawing correctly by studying it, but listening to the functioning may result in different interpretations of the message as the message is unstructured and it's context is unknown to the users. Unstructured information is dynamic (changes whenever new content is added) and therefore more prone to misinterpretation and error. Thus, structured information is more credible than unstructured information.

We extend the same logic to argue that structured *knowledge* is more credible to the users than unstructured knowledge. The differences in structured and unstructured information (Table 1) are applicable to the structured and unstructured knowledge as well. We make an assumption in this claim that the sources of knowledge in both structured and unstructured are equally credible, as long as their sources are within the organization. We do not consider sources of knowledge that are not produced or managed within the organization, such as documents obtained from the internet. In other words, we are treating the credibility of various knowledge sources as the same when that source is an individual or artifact within the organization.

Other than the differences mentioned in Table 1, we argue that the credibility of unstructured knowledge is less than that of structured knowledge for two reasons. First, unstructured knowledge is often possessed by single individuals in the organizations such as experts. For unstructured knowledge the users have to first *find the sources of the unstructured knowledge* before accessing it; second, the *transfer process of unstructured knowledge is often difficult* because of its tacit nature. Unlike accessing structured knowledge by search and retrieval mechanisms, there are no formal methods established to access unstructured knowledge. The structured knowledge can be categorized and stored in the organization and therefore easily accessible to the users. But once the unstructured knowledge is converted to the structured knowledge, the credibility of the structured knowledge increases to the users as they can now access and share the knowledge. For example, experience of an expert can be converted to explicit knowledge such as "best practices", case studies or stories. This knowledge becomes more credible to the users as it is now available in explicit form and could be easily accessible. The task of KMS that manage unstructured knowledge resources is more difficult as the knowledge cannot be accessed easily. From the above arguments we therefore claim that in most of the cases structured knowledge is *more credible* than unstructured knowledge. (It is to be noted that the present discussion is on degree of credibility – that some piece/source of knowledge is less credible than another does not mean that it will not be believed, much less that it will be as false; rather, it will be less likely to lead to belief on the part of the user.)

4.2 Veritistic gains of KMS

We claim that KMS that manage structured knowledge bring more veritistic gains to users than those that manage unstructured knowledge. KMS make available knowledge to the users, which otherwise might have been difficult to obtain. While accessing this knowledge, an individual increases his/her knowledge. Goldman (1999) suggests that the higher an individual's degree of belief in a true proposition, the more knowledge this individual possesses. In other words, an individual acquires knowledge if this individual increases his or her degree of belief in true proposition. When a user uses a KM system, he/she uses the acquired knowledge to do certain tasks that can be stated as propositions. For example, a task could be 'how to evaluate salesmen to distribute incentives?' A KM system can help in defining evaluation criteria for incentive distribution based on the past information. Users' belief for doing the task will change when he/she obtains the evaluation criteria from the KM system. We therefore view KMS as tools that modify users' belief in doing tasks. If using KMS can increase the state of belief of users from (a) false belief to true belief, (b) false belief to partially-true belief, or (c) partially-true belief to true belief then we can claim that KMS bring veritistic profits to the users.

Goldman mentions two cases where communication can yield veritistic profits. First, "when a communicated message contains a direct answer to some question that interests the receiver". And second, "when a communicated message does not contain a direct answer to a question but contains a report of some evidence that the receiver uses to answer her question" [6] (p. 164). The second case applies more for KMS usage. The users use the knowledge that the KMS manage as evidence in doing tasks.

We previously analyzed that structured knowledge is more credible than unstructured knowledge, therefore beliefs of users would be changed more positively (false to true, false to partially true or partially-true to true) using structured knowledge than using unstructured knowledge. When KMS manage structured knowledge, the change in users' belief from false to true, false to partially true or partially-true to true would likely be *high* and therefore users will be able to draw accurate conclusions using this type of KMS. Likelihood of the veritistic gains of the users will be high in this case. The situation will be reversed when users use KMS that manage unstructured knowledge where likelihood of veritistic gain is *low*. This finding can also be generalized to the practice of KM in organization. If on average (average value of individual V-value) the practice of KM increases the V-value of the user's belief status, then the over all practice of KM in the organization will bring veritistic profits to the users.

The above analysis is based on 'likelihood of veritistic gain' instead of simply 'veritistic gain' because there may be some beliefs that are irrelevant or unimportant to the users. These beliefs, though few in numbers would be viewed as uninteresting and therefore would not bring veritistic gains to the users.

We made two assumptions in this analysis. First, users need to have genuine interests on propositions or artifacts to be fit for veritistic analysis. KMS that are used in organizations help the users to take decisions in the organizations and therefore perceived to be useful. Therefore knowledge managed by KMS is of interest to the users. Second, it is assumed that the technologies used in the KMS are most appropriate and correctly chosen. The degree of trust users place in the KMS also depends on the technologies used in the KMS and we assume that users trust the technologies.

5 Conclusion

The change in beliefs of the users to make certain decisions will depend on the KMS that they use. KMS manage knowledge, and users use this knowledge to perform certain tasks within the organizations. We have argued that KMS that manage structured organizational knowledge bring more veritistic gains to their users than KMS that manage unstructured knowledge. This veritistic analysis of KMS can help users to identify KMS that they are most likely to use. *Users would like to use KMS that bring more veritistic gains to them than others*. In other words, users tend to use KMS that are seen as increasing their stock of true belief, and they tend not to use KMS which they regard as not having such veritistic benefits. Our veritistic analysis of KMS usage is consistent with the current design of KMS used in organizations as Marwick [4] (p. 814) points out, "the strongest contribution to current knowledge management solutions is made by technologies that deal largely with explicit knowledge, such as search and classification. Contributions to the formation and communication of tacit knowledge, and support for making it explicit, are currently weaker".

References

1. Alavi, M. and D. Leidner: Knowledge Management and Knowledge Management Systems: Conceptual Foundations and Research Issues, MIS Quarterly, Vol. 25, no 1, (2001)
2. Nonaka, I.: A Dynamic Theory of Organizational Knowledge Creation, Organizational Science, Vol. 5, no 1 (1994) 14-37
3. Stahl, G.: A Model of Collaborative Knowledge-Building, in Fourth International Conference of the Learning Sciences, Mahwah, NJ: Erlbaum (2000)
4. Marwick, D.: Knowledge Management Technologies, IBM Systems Journal, Vol. 40, no 4, (2001) 814-830
5. Hahn, J. and M.R. Subramani: A Framework of Knowledge Management Systems: Issues and Challenges for Theory and Practice, in 21st International Conference on Information Systems, Brisbane (2000)
6. Goldman, A.: Knowledge in a Social World, Oxford University Press, New York (1999)
7. Goldman, A.: Social Epistemology, The Stanford Encyclopedia of Philosophy, ed. E.N. Zalta, plato.stanford.edu/archives/spr2001/entries/epistemology-social/ (2001)
8. Belkin, N.J.: Information Concepts for Information Science, Journal of Documentation, Vol. 34 (1978) 55-85
9. Rogers, E.M.: Diffusion of Innovation. 4th edition, The Free Press, New York (1995)
10. Hicks, B.J., Culley S.J., Allen R.D., Mullineux G.: A Framework for The Requirements of Capturing, Storing and Reusing Information and Knowledge in Engineering Design, International Journal of Information Management, Vol. 22, (2002) 263 -280
11. McDowell, A.: Trust and Information- The Role of Trust in The Social Epistemology of Information Science, Social Epistemology, Vol. 16, no 1 (2002)
12. Foley, R.: Intellectual Trust in Oneself and Others, Cambridge University Press (2001)
13. Hardwig, J.: The Role of Trust in Knowledge, The Journal of Philosophy, Vol. 88, no 12 (1991) 693-708
14. Fogg, B.J. and H. Tseng: The Elements of Computer Credibility in Conference on Human Factors and Computing Systems, Pittsburgh, PA (1999)
15. Self, C.: Credibility. An Integrated Approach to Communication Theory and Research, ed. M. Salwen and D. Stacks, Mahwah, N.J.: Erlbaum (1996) 421-441

Grounding Knowledge of Engineering Applications in Systematic Terms

Ying Liu, Jin Yu

Computing Science Department, University of Aberdeen, Kings College,
Aberdeen AB24 3UE, Scotland, UK
{yliu, jyu}@csd.abdn.ac.uk
http://www.csd.abdn.ac.uk/~{yliu, jyu}

Abstract. In the main research of internet-computing enabled knowledge management, we use some of the most advanced research scenarios, arguing that we critically need a system approach to question where knowledge comes from. In particular, within a given engineering domain, we synthesis the problems and reveal, that the knowledge is embraced by interactions among systems, system observers, observables, engineering objects and instruments; that the complex system interactions must be dispatched into infrastructural layers based on physics-ontologies; that the ontologies must be dedicated to human and data communications. Such a synthesis would impact on knowledge technologies for solving engineering problems in scalabilities, as well as in collective vocabularies that must associate with the communication crossing the layers in the problem solving environment.

1 Introduction

A crucial object to study seems to be missing in the mainstream concept of internet computing enabled knowledge management[1] (KM): where is the knowledge? Many conceded that there must be an emphasis on tools and rules for KM that is out of step with human modes of capturing, sharing, processing and determining information in business organisations [17], [7], [27]. Moreover, Wilson concludes that KM is nonsense as a field of management consultancy practice after extensively surveyed journal literatures and reports from Accenture, Cap Gemini Ernst and Young, Deloitte and Touche, Ernst and Young, KPMG Consulting, McKinsey and Company, and PricewaterhouseCoopers [42]. But, what about KM for solving problems in an engineering domain, e.g., [38], [20], [21], [22]?

Engineering is a human effort to change or facilitate a kind of environment in order to make that environment more suitable or responsive to perceived human needs and wants. Such an effort results many kinds of physical outputs; it may define, design, develop or maintain a system. Many actors take part in engineering. One group are engineers; others are managers; still others are ones who create artifacts such as numerical models according to specifications. Much knowledge is derived from human observations, designs and experiments. They all only know what they know when they need to know it. From a computing perspective, the KM must have its meta-systems in whatever forms showing how knowledge is *grounded* from a level of engineering to a level of business organisation that manages the engineering processes. At either level, the KM research must recognise the importance of methodological inter-disciplinarity, so the research can be continued and making sense [6]. There are lengthy discussions on understandings of the notion "engineering"[2], "knowledge"[3] or (symbolic) "grounding"[4]. However, to address the problems more adequately, this paper uses some advanced research scenarios and cases, arguing that we critically need a system approach to open up the "where" question, so as to reconcile some multi-disciplinary differences to enable researchers to cross the invisible boundaries, rather than continue a research in isolation from one another. As Popper noted [23]:

---

[1] A report of over 100 pages surveyed and studied systems, products and methods that challenge the knowledge sharing [39], where the author, Dave Snowden, a director of IBM's newly created Cynefin Centre for Organisational Complexity and formerly a director of IBM's Institute for Knowledge Management, calls for the third generation of knowledge management making sense models with adaptive systems theory [35]; Also, see, the book 'The Knowledge Management Fieldbook' [3].

[2] See, e.g., [30], [9], [24], [40].

[3] See, e.g., [13], [1], [15], [4], [19], [8].

[4] See, e.g., [2], [29], [44], [12], [14].

"studies or disciplines are distinguishable by the subject matter which they investigate, [this] appears to me to be a residue from the time when one believed that a theory had to proceed from a definition of its own subject matter... But all this classification and distinction is a comparatively unimportant and superficial affair. We are not students of some subject matter but students of problems. And problems may cut right across the borders of any subject matter or discipline".

Indeed, subjects in KM are increasingly diversified; so are the objects to study. Many technologies are under developing on one hand, but on the other are piled up, not added up inter-relatedly, because making the technologies deployable by one another is not that obviouse. Once a new research program is initiated, it becomes very difficult to establish a feasible framework to orientate the research. See, e.g., the lessons [23]. The position of this paper is that, in an engineering domain, knowledge is grounded in a coherent infrastructure interacting with humans, systems, data, experiments, organised communications, objects to achieve or problems to solve, as well as tools that support all this, including computers, the web, and data networks. In the systematic terms (we here use to simplify the following discussions), knowledge is embraced by interactions[5] among systems, system observers, observables, engineering objects, and instruments; knowledge is embodied in observers primarily and embedded in layers and tiers of data and human communication infrastructures; knowledge is usable only if it flows via the infrastructures and outsourced during problem solving processes[6]. The rest of the paper is organised into three parts.

First we illustrate that there are ongoing computing research themes on the KM that are largely diversified, but equally important to engineering applications. Second, in order to identify the critical problems, we review some systematic notions. Third, by using the notions, we intend to describe how the knowledge is derived from the various interactions.

2. Is Knowledge ...

2.1... from computing technologies? ...

One of the foremost objectives for computing research advancing KM today is to provide means, on account of advantages for developing the web driven technologies, to serve the needs for knowledge pervasively in a problem-solving environment (PSE)[7]. As an investigation of a realisation proceeds to achieve this end, a given computing application often shows many facets in large scales. For example, a UK e-Science program has recently argued the case for creating new types of digital libraries for automating the process from flooded raw data to knowledge. The web services, the semantic web, ontologies, meta-data, data archiving, data mining, Grid computing and middleware, multi-agents, pervasive computing, artificial intelligence, digital libraries, etc. are all deployed [10].

2.2.... from data? ...

As a case of engineering knowledge from data, or of data mining so to speak, Rolls-Royce places itself on the knowledge technologies of the future with a multi-million pounds project with AKT (Advanced Knowledge Technologies). Another £3m grid-computing research, namely, Distributed Aircraft Maintenance Environment (Dame) project is initiated for developing technologies in two areas [5, p. 12]:

---

[5] The term "system interactions" is studied in [21] for managing software production, and, for developing decision support systems in water resources management [20].

[6] In [16], Hull even views science as a process.

[7] In [46], a PSE is described as a computer system that provides all the computational facilities needed to solve a target class of problems. These features include advanced solution methods, automatic and semiautomatic selection of solution methods, and ways to easily incorporate novel solution methods. Moreover, PSEs use the language of the target class of problems, so users can run them without specialized knowledge of the underlying computer hardware or software. By exploiting modern technologies such as interactive colour graphics, powerful processors, and networks of specialized services, PSEs can track extended problem-solving tasks and allow users to review them easily. Overall, they create a framework that is all things to all people: they solve simple or complex problems, support rapid prototyping or detailed analysis, and can be used in introductory education or at the frontiers of science.

a) collecting real-time engine diagnostics while the plain is in flight and analysing the vast amounts of data gathered from the thousands of Roll-Royce engines used around the world, and
b) finding out the last time an engine went "bump" and "squeak" when a field in the database called "bump" or "squeak" cannot be found. What should also be found is all the conditions before and after what happened on the engine.

2.3.... from human?

Human problem solving is done within a context which constrains the solution space. It is human who needs the knowledge to solve a problem in that context, and knows what is known by means of data analysis, experiences, perceptions, communication, experimentation, reasoning or other kinds of cognitive processes. Considering the Rolls-Royce case mentioned above, we have a relative environment and an absolute environment. If a system to be maintained is as it is from our own knowing perspective, we are in a relative environment. If a system to be maintained is as it is in itself, we are in an absolute environment. Maintenance engineers are constantly involved in both environments. Thus, the two environments are coherent with the engineers.

3 Back to the Basics: Some Concepts, Principles and Notions

But before coming to the position that knowledge is grounded in a coherent infrastructure of all the interactions on the dimensions in section 2, we need to review some systematic concepts, principles and notions[8].

3.1 Some System Concepts and Principles

System methodology has been widely employed in solving many kinds of problems: from the concern for classic engineering analysis, such as flow of matter and/or energy, to the concern for modern-day controlling communications, such as information. Hence, for a research strategy, it forces one to look at a problem in its entirety. To contribute KM research in particular, it interrelates a large range of engineering processes which appear to be derived from different domains on one hand, but on the other, have to be constricted under different disciplines for the same objectives, and to be managed within one frame – a real or abstract simplification of the problems situated in a problem solving environment.

> *Def 1: A system is a collection of components, also called parts, either physical or non-physical in nature, that a) exhibit a set of interrelations among themselves and interact together towards on or more goals; b) exhibit properties processed differently from the collection of properties processes by the individual parts.*

Let us now consider a system in terms of engineering. In other words, a system is now considered as an object to engineer. Thus, a consideration of a system is encompassed by the most important objects or classes of objects that an engineering effort targets, including, for example,
a) a product that is to be delivered to users after completion of the necessary engineering processes;
b) the engineering processes that have to shape the product itself, such as establishments or requirements' elicitation.

> *Def 2: A system observer is someone who starts with something for some reason of his own intentions to describe that ?something? holistically, that is to say, in terms of whole elements linked in hierarchies.*

[8] Cybernetics as a field of system research has thoughly come into studying the nature of knowledge. See a good web site http://www.pangaro.com/published/cyber-macmillan.html. Here we only review the system notions for further practice-aspiration.

But, a system must be observable. For example, if there were no data coming from an engine, there would be nothing to tell about the state of that engine.

> *Def 3: A system observable is a piece of information that a system observer perceives and believes that it tells something about the system that the observer observes.*
> *Def 4: An observing instrument is anything by which that a system observer is aided to obtain a system's observable.*

From Def 1-4, we have the following principles:

> *Principle 1: A system has no existence independent of its system observer*
> *Principle 2: A system observable must exist between a system and its system observer.*
> *Principle3: In order to "see" a system, a system observer of that system often needs an instrument.*

Now let us briefly review a few technological notions.

3.2 The Web Ontology

The notion ontology has a long history in philosophy, where ontology is about a systematic account of beings, or existence of things, or to being in the abstract as a "reality". Science has shown a reality that is structured all the way down. At bottom it consists not of four types of gunk: earth, water, air, and fire, but rather of a finite number of definite particles, lawfully related one to the other [45]. Then, it comes to the hope that we might build up substantial information about our world from the elementary information we find at the bottom of reality. So, the ontology in terms of philosophy is a theory of arguing and explaining. The term 'ontology' has been used in this way for a number of years by the artificial intelligence and knowledge representation community, but is now becoming part of the standard terminology of a much wide community including object modelling and XML [27]. The key ingredients that make up the web ontology are vocabularies of basic terms and a precise specification of what those terms mean. Numerous researchers believe that the web ontology can be the useful tools for the following reasons (revised from [11]):

a) The web ontology is more than an agreed vocabulary. It provides a set of well-founded constructs that can be leveraged to build meaningful higher level knowledge. The terms in ontology are selected with great care, ensuring that the most basic (abstract) foundational concepts and distinctions are defined and specified. The terms chosen form a complete set, whose relationship one to another is defined using formal techniques. It is these formally defined relationships that provide the semantic basis for the terminology chosen.
b) The web ontology is more than a taxonomy or classification of terms. Although taxonomy contributes to the semantics of a term in a vocabulary, the web ontology includes richer relationships between terms. It is these rich relationships that enable the expression of domain-specific knowledge, without the need to include domain-specific terms.

The last notion we need to explain here is "infrastructure".

3.3 The Infrastructure

Infrastructure seems to be singularly boring as an object for scientists to study. It is often referred to as a list of technical specifications, black boxes, places, wires, plugs, roads, bridges, stations, etc. Infrastructuring is usually seen as an engineering work to establish public services and utilities for social communities. Roads, railways, bridges, pipe lines, electricity, etc. are instances of social infrastructures. Because of the world's technical sound, people now use the term infrastructure to refer to any substructure or underlying structure of systems [36] - most notably the information superhighway - the global information and communication infrastructure of networks that include the

Internet, WWW, telephone networks, cable, satellite, wireless, or electronic sensor based data networks. These are the backbone infrastructures of our electronic communications today. The web is a collection of interlinked electronic items including documents, texts, images, music files, video, etc. hosted on servers all over the world, mostly hosted on HTTP (Hyper Transfer Protocol) servers [33]. The web lives on the internet by a set of protocols running over the net. Although the web is part of the net, the net is much larger than the web. The net hosts e-mail, FTP, peer-to-peer, VPNs, telephony, etc. So, there are software coded layers and tiers driven by the web: servers, clients, peers, portals, gateways, or protocols as many as the technical approaches and tools to design them, most notably, XML/RDF, the metadata, the semantic web, the web service, and the web ontology.

4 Knowledge Grounding

In this section, we use the notions defined to describe where the knowledge comes from. Again, we use the Rolls-Royce case in section 3.2. Reportedly, the case is significant [5], because some 44 per cent of Rolls-Royce's revenue comes from the maintenance and servicing of its engines in aircraft, ships and power stations. Instead of selling engines to airlines, the firm charges for use of the trusted they provide, on a "power by the hour" basis. The sooner engineers can be made aware of problems, the quicker they can be resolved and the longer a plane can spend in the air, earning the money for the supplier. Thus, the advent of the web driven KM technologies has presented huge opportunities for the operational improvements. The foremost objective in achieving this is to provide the latest and evidential information indicating the system's conditions, so to improve the efficiency of the operational processes taking place pervasively and, in turn, to have significant cuts in the cost of the maintenance. However, this strategic vision needs to have a more careful justification for research into more dedicated infrastructures.

4.1 In Physics Based Infrastructure

At this layer, the purpose of system diagnostics or monitoring is to reduce actions taken on the basis of judgements made from directly measured or inferentially calculated information with varying degrees of emphasis dictated by the needs and capabilities of their individual organisations. Engineers are interested in both determination of initially installed system's condition and in condition throughout the operation phase.

Intrinsically, the maintenance processes are rarely repetitive in the same manner as normal operational tasks. The processes do not lend themselves to systemisation and computerisation. Specifically, in an absolute environment referred in section 2.3, advances in sensor technologies, instrumentation, microprocessor-based controllers are increasingly used [25], [41]. Networks of sensory or actuator nodes with computational capabilities, connected wirelessly or by wires are getting much cheaper. Components are increasingly complex in structures and functions, but becoming more and more reliable. Only the down-time experienced can be very large. Taking all this into an account, engineers often face unknown or known problems that require various skills, kinds of information and knowledge through the interactions among (see, Def 1 – Def 5):

a) systems to be observed: an engine turbine and its properties as such time, functions, conditions, or states;
b) observers: engineers; engineering objects, acceptable engine conditions;
c) instruments: sensors, test appliances, computers, data process software, engine monitors;
d) observables: data artefacts, time series data patterns.

Let us assume there are web ontologies in this type of physics based infrastructure; we call the ontologies "physics -ontologies". We then have a layer of physics-ontology-infrastructure.

4.2 In Physics-ontology-infrastructure

By the notion "web ontology", the key ingredients that make up ontology are vocabularies of basic terms and a precise specification of what those terms mean. But abstracting the vocabularies and their relationships is only one way to support human communications. To make the ontology operational on persistent tiers and layers of infrastructures is something entirely different. It is on these layers where data are collected, distributed and measured; that reports are circulated; and that groups are participating and communicating with one another. Data in a physics based infrastructure cannot be explained merely as a consequence of a differing coherence of an utterance. They depend on who makes the utterance, where the sensors are situated, where the data are channelled, how the data are stored and filtered, or what methods are used to understand and explain an observed phenomena.

Thus, the web ontology must be systematically constrained by the physics based infrastructure as Kharkov has also studied [18]. The priori knowledge for the ontology design must be closely inherent to understandings of physical systems, as well as practical experience with the systems. A problem solving process for a given application can then be supported by the "content" of the priori system information. The third interactive layer is human oriented. We note this as human-physics-ontology-human communication infrastructure.

4.3 In human-physics-ontology-human communication infrastructure

At a level of management, the system maintenance is extremely critical for industrial companies to sustain their productivity. An engine's maintenance is no longer just a traditional event of a repair – call an engineer in with parts and tools to fix it. It is a matter of how to detect the first sign from the engine, so something is known priorily if there is a need for preventing the "disasters". This is the essential idea behind the method called condition based maintenance [43].

Whether to proceed with the emergency repair may well be informed by the effects on the bottom line in the physics-based infrastructure. Maintenance can be based on equipment run times and starts and stops thus providing the basis for predictable maintenance. Engineers can properly analyze equipment failures and forecast the probability of the same equipment failing in the same plant or other business units, or undertake the processes, such as data collection, data clustering, testing, fault or defect diagnosis, planning spare parts, making recommendations, reporting major factors affecting a system' s life, all in a technical and timely manner.

All the web layers are meaningful and usable only when a system observer participants in a particular communication [34]. Whether a maintenance engineer can exploit in elliptical or anaphoric resolution is depending in part on the role that the engineer has most recently played in the communication in the physics-based infrastructure. As Quine remarkably observed and his points are still significantly relevant for today [32]:

> "the things in sharpest focus are the things that are public enough to be talked of publicly, …., and near enough to sense to be quickly identified and learned by name and labels; Moreover, a common sense talk of physical things often goes forward without benefit of explanations in more intimately experimental terms. …If we improve our understanding of ordinary talk of physical things, it will not be by reducing that talk to a more familiar idiom; There is none. It will be by clarifying the connections, causal or otherwise, between ordinary talk of physical things and various further matters which in turn we grasp with help of ordinary talk of physical things" .

5 Conclusion

To reconcile some multi-disciplinary differences and to cross the invisible research boundaries in KM research, system methodologies enable us to dispatch the complex knowledge grounding contextures into hybrid infrastructural layers. In an engineering domain, knowledge is embraced by interactions among system observers, systems, observables, engineering objects, and instruments. Knowledge is embodied in observers primarily, embedded in layers and tiers of infrastructures, and only usable if it flows via the infrastructures and outsourced during a problem solving process. KM requires such an infrastructure as a higher order system environment to be understood and controlled, so as to dedicate KM methods, solutions and practice to a problem solving process. Such a system synthesis is crucial for KM; it considerably impacts on a) a system scalability for a given application, b) a level and a scope of ontology design, and c) collective vocabularies in ontology design that must associate with meanings to be understood within the infrastructures.

References

1. Benrnecker, S & Dretske, F.: Knowledge – Readings in Contemporary Epistemology. Oxford University Press (2000)
2. Bryant, A.: Grounding Systems Research: Re-establishing Grounded Theory. In: Proceedings of the 35th Hawaii International Conference on System Science, IEEE Computer Society Press (2002)
3. Bukowitz, W. R., Williams, R. L.: The Knowledge Management Fieldbook. Financial Times-Prentice Hall (2001)
4. Compbell, D. T.: From Evolutionary Epistemology via Selection Theory to a Sociology of Scientific Validity. *Evolution and Cognition (1997)* 3, p. 5-38
5. Computing: Rolls-Royce Is Confident Grid Technology Will Earn its Wings. Written by Bryan Glick, *Computing*, 30 April (2003)
6. Dervin, B.: Human Studies and User Studies: a Call for Methodological Inter-disciplinarity. *Information Research*, Vol. 9 No. 1, October (2003)
7. [Does KM=IT? Hompage], http://www.brint.com/advisor/a091599.htm
8. Dretske, F.: Knowledge and the Flow of Information. Basil Blackwell (1981)
9. Dym, C.: Engineering Design: A Synthesis of Views. New York: Cambridge University Press (1998) p. 15
10. eSI-Report: Scientific Data, Mining, Integration, and Visulization. In: Mann, B., Williams, R., Atkinson, M., Brodlie, K, Storkey, A. & Williams, C.: Report of the Workshop Held at the e-Science Institute. Edinburgh, 24-25 October (2002) http://www.cacr.caltech.edu/~roy/papers/sdmiv-ltr.pdf
11. Fensel, D.: Ontologies: Silver Bullet for Knowledge Management and Electronic Commerce. SpringerVerlag (2001)
12. Galison, P.: Contexts and Constraints. In: Buchwald, J. Z. (eds.): Scientific Practice: Theories and Stories of Doing Physics. Chicago, University of Chicago Press (1995)
13. Gennari, J. H., Musen, M. A., Fergerson, R. W., Grosso, W. E., Crubezy, M., Eriksson, H., Noy, N. F., Tu, S., W.: The Evolution of Pretege: an Environment for Knowledge-based Systems Development. *Int. J. Human-Computer Studies,* 58 (2003) 89-123
14. Harnad, S.: The Symbol Grounding Problem. *Physica*, D 42 (1990) 335-346
15. Hooker, C. A., Christensen, W. D.: The Organisation of Knowing: Autonomy, Evolution, and Evolutionary Epistemology. In: Festschrift for Donald T. Campbell, Hays, C. & Hull D (eds.). Chicago: University of Chicago Press (1998)
16. Hull, D.: Science as a Process. Chicago, University of Chicago Press (1988)
17. IBM: Next Generation Knowledge Management: The complexity of Humans. Executive Tek Report, *IBM Global Services*, 13 November (2002) http://www.ibm.com/services/files/etr_cynefin.pdf
18. Kharkov: Ontology Development for Gas Compressing Equipment Diagnostics Realized by Neural Networks. Report on System Development Project, Kharkov's Branch, Artifical Intelligence Department, Industrial Ontologies Group, The Ministry of Education and Science of Ukraine, Kharkov National University of Radioelectronics, Ukraine 2003) http://www.cs.jyu.fi/ai/OntoGroup/NN_Ontologies.pdf
19. Levinson, P.: Mind and Large: Knowledge in the Technological Age. Jai Press INC (1998)
20. Liu, Y., Veer, vd.P.: To Approaching an Intelligent System in Coping with an Aspect of Complexity. In: Proc. of the 1993 ACM Symposium on Applied Computing, Indianapolis (1993)
21. Liu, Y.: Treating the Interactions Systematically. *Software Engineering J.*, IEE/BCS Press, March (1994)
22. Liu, Y.: Decision Support Environment for Engineering Use", In: Blain, W. R., Katsifarakis, K. L. (eds.): Water Resources and Distribution. Computational Mechanics Publications (1994)
23. Liu, Y.: Geo-referenced Digital Libraries: Experienced Problems of Purpose and Infrastructure. *Library Philosophy and Practice*, Vol. 6, No. 1, Fall (2003) http://www.webpages.uidaho.edu/~mbolin/liu.pdf

24. Mitcham, C.: Thinking through Technology: The Path between Engineering and Philosophy", Chicago: University of Chicago Press (1994) p. 19-38
25. Muruganandam, T. M., Kim, B., Olsen, R., Patel, M., Romig, Seitzman, J. M.: Chemiluminescence Based Sensors for Turbine Engines. In: 39th AIAA/ASME/SAE/ASEE Joint Propulsion Conference & Exhibit, AIAA 2003-4490, Huntsville, Alabama, USA, 20-23, July (2003)
26. Nalhotra, Y., Is Knowledge Management Really on Oxymoron? Unravelling the Role of Organizational Controls in Knowledge Management. In: White, D. (ed.): Knowledge Mapping and Management. Hershey, PA: Idea Group Publishing (2002) p. 1-13 http://www.brint.org/KMOxymoron.pdf
27. Nambiar, U., Bressan, S., Lee, M. L., Li, Y-G.: Current Approaches to XML Management. *IEEE Internet Computing*, July-August (2002)
28. Vesilind, P. A., Alastair, S.: Gunn, Engineering, Ethics, and the Environment. New York: Cambridge University Press (1998)
29. Paek, T.: Grounding Criterion: Toward a Formal Theory of Grounding. Technical Report, MSR-TR-2000-40, Microsoft Research, Microsoft Corporation, Redmond, WA 98052, USA (2000)
30. Podgorecki, A., Alexander, J., Shields, R. (eds.): Social engineering. Carleton University Press (1996)
31. Popper, K.: Conjectures and Refutations – The Growth of Scientific Knowledge. London and Henley, Routledge and Kengan Paul (1963)
32. Quine, W.v. O.: Word and Object. MIT Press (1960)
33. Schatz, B. R., Hardin, J. B.: NCSA Mosaic and the World Wide Web: Global Hypermedia Protocols for the Internet", *Science Magazine*, Vol. 265, 12 August (1994) 895-901
34. Siebes, R.: Peer to Peer Solutions in the Semantic Web Context: An Overview", EU-IST Project Report, IST-2001-34103 SWAP, Vrije Universiteit Amsterdam, Holland (2002)
35. Snowden, D.: Complex Acts of Knowledge: Paradox and Descriptive Self-awareness. *J. of Knowledge Management*, Vol 6, No. 2. (2002) http://research.microsoft.com/users/timpaek/Papers/MSR-TR-2000-40.pdf
36. Star, S. L.: Got infrastructure? How Standards, Categories and Other Aspects of Infrastructure Influence Communication: In: *the 2nd Social Study of IT Workshop at the LSE ICT and Globalization*, London, 22-23 April
37. Yu, J., Hunter, J., Reiter, E., Sripada, S.: An Approach to Generating Summaries of Time Series Data in the Gas Turbine Domain", In: Proceeding of ICII2001 (Beijing), IEEE Press, p. 44-51, Received 'Outstanding Paper' award at ICII2001 (2001)
38. Yu, J., Reiter, E., Hunter, J. & Sripada, S., SumTime-turbine: A knowledge-based System to Communicate Time Series Data in the Gas Turbine Domain. In: IEA/AIE-2003 (The 16th International Conference on Industrial & Engineering Applications of Artificial Intelligence and Expert Systems) (2003)
39. White-Paper (2002): The Knowledge Sharing Challenge. The Sensemaking White Paper. http://www2.parc.com/istl/groups/hdi/papers/ sensemaking-whitepaper.pdf
40. Wiener, N. (1989)[1950], The Human Use of Human Beings. London, Free Association Books (1989)[1950]
41. Wild, P.: Industrial Sensors and Applications for Condition Monitoring", London: Mech. Eng. Publishing (1994)
42. Wilson, T. D.: The Nonesense of Knowledge Management. *Information Research*, Vol. 8, No. 1. October (2002)
43. Wireman, T.: Computerized Maintenance Management Systems. Industrial Press, Inc (2003)
44. Ziemke, T.: Re-thinking Grounding. In: Riegler, A.., Peschi, M (eds): Does Representation Need Reality? – Proceedings of the international conference 'New trends in cognitive science (NTCS 97), (1997) p. 87-94
45. Zimmerman, D. W.: Could Extended Objects be Made Out of sSmple Parts? An argument for 'atomless gunk", *Philosophy and Phenomenological Research*, 51 (1996) 1-29
46. Houstis, E., Gallopoulos, E., Bramley, R., & Rice, J. (eds., Problem-solving Environments for Computational Sicence. *IEEE Computational Science & Engineering,* special issue, July/September (1997)

Logic, Neuroscience and Phenomenology: In Cahoots?

Ahti-Veikko Pietarinen*

Department of Philosophy, P.O. Box 9, FIN-00014 University of Helsinki
pietarin@cc.helsinki.fi

Abstract. Cognitive sciences, including cognitive neurosciences, have provided important insights into the notionsof awareness, implicit/explicit information processing in knowledge, perception, object identification and memory, as well as general information retrieval. Meanwhile, propositional-attitude logics have tried to account for awareness in terms of symbolic tools, but have not found pathways in which to relate the two fields. It is argued that empirical findings concerning rare neural dysfunctions (blindsight, unilateral neglect, prosopagnosia, implicit memory) may contribute to these logical investigations. On the other hand, the early phase on cognitive science, the origins of which coincide with that of pragmatist philosophy, shared intimate roots with phenomenology. Accordingly, I will identify some strands in that early period that have surfaced in logic, AI and computer science. In phenomenology, the significance of the division between implicit and explicit aspects of knowledge in understanding cognition was acknowledged very early on.

1 Introduction

One of the key conceptual tools in cognitive neurosciences is the *implicit/explicit* distinction. It may be drawn in similar ways with respect to a variety of notions such as knowledge, belief, perception, memory and learning. These notions have been, in the mainstream analytic philosophy, subsumed under the concept of *propositional attitudes*, especially epistemic ones. Unfortunately, this perspective masks the processual, active and dynamic character of these notions. Above all, it masks the difference between implicit and explicit methods of knowing, believing, seeing, recalling or learning, a key descriptive division in cognitive approaches to information processing in the human brain.

Empirical and *logical* sides of these manifestly different implicit/explicit distinctions may nevertheless be examined in a parallel fashion [20]. The goal is to provide cognitive neuroscientists, computer scientists and philosophers with integratory tools that smooth progress in mutual understanding of what it means for the mind to be simultaneously both conscious and an aware, and unconscious and unaware processor of information. This amalgamation is expected to provide

* Supported by the Academy of Finland (Project No. 103130). My thanks to the reviewers for comments.

new cognitively-grounded logical tools for both philosophical and computational purposes, and contribute to the drafting of a preliminary agenda of informatics that would not evolve in isolation from neighbouring disciplines.

In this paper, I will delineate some parallels between logical and neuroscientific aspects of awareness and the implicit vs. explicit distinction, with special reference to *phenomenology.*

2 Relating Neuroscience and Logic

2.1 Implicit versus Explicit Knowledge

It is quite striking to become aware of the extent to which neuroscientific research contributes to conceptual and logical approaches to the epistemic concept of knowledge and its ilk. Indeed, in recent years, cognitive neuroscience has provided insights into implicit versus explicit information processing in *perception, identification, memory, information retrieval, belief content, belief formation, iterated knowledge* and *introspection*, to name but a few [6, 27]. Perhaps above all, cognitive neuroscience has brought the notion of awareness to the forefront of human knowledge [28].

Meanwhile, philosophers and cognitive scientists have suggested diverse approaches to problems posed by the mental concept of *awareness* [5]. There are significant logical reflections of neuroscientific phenomena that have remained uncharted. But neuroscientific insights may be carried over to bear on logical theories. A particularly useful way of doing this is by introducing *operators* into epistemic languages aimed at formalising constructions of agent's knowledge and other variants such as memory and belief [8, 10, 16, 17, 20].

This has some inevitable repercussions to general theories addressing the *logic of consciousness.* To date, however, logical theories that aim at incorporating some interpretation of the notion of awareness into the language have merely been addressing the question of how to dispose of the *logical omniscience problem* in the resulting systems. That is to say, they have been baffled by the property of traditional epistemic logics, namely that agents come to know all the logical consequences of known formulas. While the problem of logical omniscience may have been the first job description of the introduction of the notion of awareness to bear on logic [8], this no longer needs to be the case. Logical omniscience may be extirpated by other, much stronger means, for example by augmenting the received *possible-worlds semantics* for knowledge with *impossible possible worlds*, in which not all the classically-valid theorems, such as the law of non-contradiction, hold in the sense of these worlds being epistemically possible — even if they were to accommodate contradictions [24]. In addition, it has been shown that logics of awareness may be embedded in the impossible worlds framework [26]. Because (but not only because) of this fact, vital applications for the task of bringing logical concerns closer to cognitive ones are better to be sought elsewhere than in the omniscience problem.

This said, the logical omniscience problem is not devoid of explanatory power with respect to neuroscientific issues. It provides a logical answer to the neuroscientific problem of what it is that separates the implicit and explicit aspects of knowing from one another. The customary explanation is that implicit knowledge does not exhibit *conscious access to information.* From the logical viewpoint, it may be said that an agent implicitly knows something iff that knowledge is *closed under a logical consequence relation.* The phrase 'conscious access to information' recently ingerminated in neuroscientific literature turns out to carry logical content.

Conversely, cognitive neuroscience puts before us a range of data concerning *blindsight, unilateral neglect, prosopagnosia* or *implicit memory* not prehended in logic yet. In fact, a wealth of conceptual and interpretational issues exists to be assessed for relevance of neuroscience to logical modelling and the analysis of knowledge and cognition in relation to awareness.

A genuine need exists for a unifying language in which these investigations may be carried out. One such candidate is epistemic logic, the *epistemic variant of modal logic* dealing with an agent's propositional attitudes, introduced in [10]. Indeed, both epistemic logic and the associated possible-worlds semantics turn out to be quite versatile in modelling actual cognitive phenomena (cf. some related suggestions to that effect in [17]). These phenomena include the distinction between explicit and implicit knowledge, belief and memory, different senses of awareness, and perception. Accordingly, new ways of extending the basic epistemic language to cover empirically-supported phenomena in cognitive neuroscience such as blindsight and aspects of amnesia and memory become amenable to a new kind of conceptual analysis. The relevant cognitive phenomena are, in fact, quite overwhelming: [20] parallels only a few logical and neuroscientific facets of the implicit vs. explicit distinction in relation to knowledge, memory, and other propositional attitudes, and there are certainly more.

Hintikka [11] was perhaps the first to recognise the importance of the interrelations between experimental findings in neuroscience on the one hand, and epistemic logic on the other. He discussed these relations with an eye on philosophical insights into the famous cogito argument 'Cogito, ergo sum'. One of the primary insights in that paper was the differentiation between two modes of *identification*, the *perspective* and the *public* one, and to relate that distinction to neuroscientific and cognitive 'where-versus-what' systems that have independently been observed as relevant to higher-level cognitive functions.

2.2 What is the Logic of Awareness?

Without probing into the details of the approach [17, 20], in *logics of awareness*, an agent i can be said to be aware of a proposition p in case in which i knows that $K_i\,(p \vee \neg p)$. In other words, the agent knows that the law of excluded middle holds. This has sometimes been explicated in the sense of *situation semantics*, so that the agent is aware of the proposition precisely in those situations that 'support' or 'preserve' the truth-value of the proposition. In other words, an agent cannot be aware of propositions that have a truth-value of Undefined. In alternative terminology, there are no partial interpretations in such situations for sentences among the agent's set of 'aware' ones.

In [8], awareness is taken to be a syntactic operator, conjoined to an implicit knowledge operator precisely in order to convert the proposition into explicit knowledge. (A related distinction exists in neuroscience, in which one would instead speak of *overt* vs. *covert* knowledge [27, p. 256].) In other words, given implicit knowledge $\mathbf{K}_i\,\varphi$, this is transformed into explicit knowledge by defining $K_i\,\varphi ::= \mathbf{K}_i\,\varphi \wedge A_i\,\varphi$. The operator A_i attached to the proposition φ means that i *is aware of* φ, and is explicated in [8] by the syntactic means of denoting an 'aware' proposition. To this effect, a mapping $\mathcal{A}_i(w)$ from a world w to propositions provides a set of propositions of which agent i is aware.

The suggestion that as an operator awareness may be interpreted in different ways depending on the purpose at hand is not really motivated by conceptual or cognitive concerns, but by *computational* ones. It is contestable whether computation is a likely candidate for a general theory of cognition. Such readings are proposed in [8] as 'an agent is aware of p if she is aware of all concepts that it contains', or that 'p's truth-value can be computed within an interval T'. There are other conceptual, cognitive and computational readings of awareness, yielding to independent logical systems of their own [20].

2.3 Three Levels of Investigation

The cognitive workings of awareness and related areas of information processing in the brain may be studied at several levels. The first is the actual, *physiological* and *neural end* of mechanisms of information processing. This level is intimately connected with the details of how to *implement* a certain kind of awareness in a particular hardware configuration, irrespective of whether such hardware consists of neural processors, bio-computers, quantum gate devices, or any of the platforms employed in traditional computing.

The second level looks away from these actual mechanisms and examines those tasks in which an understanding of awareness is sought via *cognitive information processing*. Examples of this *task-oriented level* are problems in *psycholinguistics* such as text or dialogue processing and comprehension, or the role of the cross-modal (perceptual, tactile, auditory) conception of awareness in abstract skills such as problem-solving and reasoning. Memory and information retrieval are also relevant in these tasks.

On the third level, we look away from both actual neural mechanisms and task-oriented analysis and develop *abstract* (say, *logical* or *semantic*) ways of addressing awareness. It is this third level that, in my opinion, sets the most topical agenda for the interplay between philosophy and informatics. This methodological stance is not intended to diminish the role of computation *per se*, but that concerns only a derivative agenda, materialised once the logical workings of cognitive notions are better understood.

This is by no means the only viable standpoint to the overall task of marrying philosophy, cognitive and computational sciences with one another. At the very least, complex co-acts between the three main levels and their sub-levels appear incontestable in any broad-minded cognitive theory. But if so, then the input provided by neuroscience is not to be ignored on the other two levels, either, no matter how abstract (e.g. non-empirical) the level of analysis may be.

The interplay between logical notions of knowledge, belief, perception and memory on the one hand, while setting cognitive neuroscience and neuropsychology on the other, also raises its head in artificial systems. For example, in *systems controlled by* such languages that involve epistemic notions there is the need to create proper naming relations instead of just generic terms in order to attain tasks that involve the conscious processing of information. Logical systems are illustrations of what is generally required of such epistemic languages, and they are for that reason useful for several *knowledge-representation tasks* concerning *multi-agent systems* [22].

However, this does not have to lead to any *reduction*, for in viewing either side of what is much the same matter. Both logical and neuroscientific terms and concepts are needed when approaching intelligence in cognitive systems in order to embed more reliable and realistic epistemological features into such artefacts. There is little substance in those arguments that claim that one day, by looking at subneuronal levels in the brain, we may find logical operations and properties in action at those levels.

2.4 Binding as a Trans-Disciplinary Problem

Although the logical systems of implicit knowing and attitudes outlined in [20] are propositional, the addition of quantifiers would in the end be indispensable, since full notions of knowledge cannot exist independently of how objects are identified, a fact that is intimately connected with the neuroscientific phenomena of how humans actually individuate objects [17]. However, such extensions give rise to new questions: Is *cross-identification* viable between different compartments of sets of possible worlds needed in disgregating viable implicit acts or attitudes from explicit ones? If so, will cross-identification be related to the neuroscientific *binding problem*, namely the question of how do independent bits of data come to be combined into unitary, coherent percepts?

There is a wider insight motivating the topicality of these questions. In logic as well as in neuroscience, *binding data* is vital. Symbolic, logical systems typically achieve this by reusing the same variables in a formula, in other words arranging variables to become bound by the same quantifiers. In neuroscientific terms, the problem has been to formulate a conjunction of different local areas of the brain capable of producing *coherent experiences*. The traditional method in logic is to rewrite universal (resp. existential) quantifications as infinite conjunctions (resp. disjunctions) of atomic predicates. On the other hand, a richer (albeit historically earlier) way is to consider both the *processes of predication* and *identity* between different occurrences of variables as reflections of the same underlying conceptual process. The nature of this unifying process may logically be unravelled as a diagrammatic and iconic rather than a symbolic activity. One fallout is that by such a *diagrammatisation*, an alternative and formally rigorous path to conceptualisation processes advocated in *cognitive linguistics* [14] is revoked.

More generally, the affinity between binding data and binding percepts vindicates Lakoff & Johnson's [14] conclusion that "to understand reason we must understand the details of out visual system, our motor system, and the general mechanism of neural binding" (p. 4). In diagrammatic approaches to logic [18, 21], predication, existence and identity are attained via the same underlying sign that marks continuous connections between differently localised predicates. I believe that such *heterogeneous, iconic* and *topological representations* will provide a promising locus for one of the most striking illustrations of the congeniality of two hitherto-isolated areas, logic and cognitive neuroscience.

3 Phenomenological Ramifications

3.1 Phenomenology: an Aide to Artificial Intelligence after all?

Phenomenology has traditionally been inimical to AI [7]. However, Albertazzi [1] has recently revealed a rich conceptual repertoire and thematic complexity in the early Central European theories revolving around the accounts of cognitive science developed in between 1870 and 1930. They include psychophysical, functional and constructivist approaches to cognition. These theories, she argues, anticipated much of the content of the contemporary cognitive sciences as well as many of the modes of research typical of the current mainstream research in cognitively-oriented sciences and philosophy.

Let me add that the revival of this early endeavour to develop *conceptual categories of the mind and psyche* is currently underway in several junctures of AI, logic and cognitive science. Unfortunately, this historical connection has not featured ofttimes in these theories. The following points need to be highlighted thus:

- Scientists and logicians, who instead of *objects* were accustomed to talk of *individuals* during the symbolic era, have been resorting to the more phenomenological terminology.
- In lieu of *predicates*, many logical theories, especially the computational ones dealing with issues revolving around 'practical reasoning in rational agenthood' [9] have switched to pictorial, visual, graphical and iconic modes of representation in order to capture different notions of *qualities* associated with assertions and objects [21].
- A related tendency is to dispense with the traditional notion of *logical constants* and substitute it with an assortment of iconic, topological and similar *Gestaltpsychologische* notions in *dynamic theories of action and experimentation* concerning the *relations* involved in the representations (e.g. conceptual graphs in AI). Similar dissociations are manifest in tendencies to dismiss the division between logical/non-logical constants as any good logical counterpart to the analytic/synthetic division.

Largely as a result of the recent investment in the foundations of computational sciences, the early European cognitive theories are thus, however unintentionally, beginning to be discernible in the concepts of *embodied and enactive minds* [25], *game-theoretic and open-systems approaches* to verification, refinement and composition of concurrent processes and programs [4], *reactive and embedded* computational systems in robotics and AI [15], *interactive, emergent* and *synthetic notions of meaning* entertained in cognitive semantics [14], and in *evolutionary approaches to linguistic meaning, semantics and pragmatics* [23].

This revival is, I believe, due to two principal factors: the decline of the mainstream *logico-formal* and *analytic paradigm* that dominated the better part of the twentieth-century philosophy, and the revitalisation of the *pragmatist stance* in logical philosophy, especially the Peircean one, the origins of which coincide with the origins of the early European contributors to cognitive science.

3.2 Phenomenology and Implicit Knowledge

Husserl's concept of *noema* [7, 12] has predominantly been interpreted transformationally, according to which *unconscious inferences* turn *sense data* into *perceptions*. Alternatively, it has been argued that in noema, mental activity plays a vital role in determining what the object types are that agents intentionally choose among the alternatives presented to the mind [2]. *Representational content* is, under this view, a complex of precepts via which other perspectives are synthesised and delineated.

Accordingly, one is tempted to think of there being a correlation between, on the one hand, *implicit and explicit aspects of the representational content* to which a subject's mind is attuned to, and on the other, the *character of the kind of mental activity* that determines which object types are selected to become conscious, observational and articulated knowledge. In the very least, such a phenomenological ramification resonates well with the view of consciousness in Peirce's *pragmatist philosophy* and his phenomenology as *phaneroscopy*, developed slightly earlier than Husserl's:

> Thus, all knowledge comes to us by observation, part of it forced upon us from without from Nature's mind and part coming from the depths of that inward aspect of mind, which we egoistically call *ours*; though in truth it is we who float upon its surface and belong to it more than it belongs to us. Nor can we affirm that the inwardly seen mind is altogether independent of the outward mind which is its Creator. (*Collected Papers of C.S. Peirce*, 7.558, c.1893.)

Observations, however implicit or explicit, are both *inward* and *outward*. It would be tempting to learn what light cognitive neuroscience can throw on *that* division.

4 Conclusions

Some initial, overlapping vocabularies in the interfaces of logic, cognitive neuroscience and phenomenology were delineated. Precisely how intense these tentative connections will be, or how concrete the practical relevance of these speculations (aside from the philosophy of informatics) turn out to be, must be left for future occasions to decide.

References

1. Albertazzi, L. 2001a. Back to the origins, in [3, pp. 1–27].
2. Albertazzi, L. 2001b. The primitives of representation, in [3, pp. 29–60].
3. Albertazzi, L. (ed.), 2001c. *The Dawn of Cognitive Science*, Dordrecht: Kluwer.
4. de Alfaro, L. 2003. Game models for open systems, *Proceedings of the International Symposium on the Verification Theory in Practice*, Springer.
5. Churchland, P.M. 1993. *Matter and Consciousness*, Cambridge, MA: MIT Press.
6. Dienes, Z., Perner, J. 1999. A theory of implicit and explicit knowledge, *Behavioural and Brain Sciences* 22, 735–755.
7. Dreyfus, H.L. (ed.), 1982. *Husserl, Intentionality and Cognitive Science*, Cambridge, MA: MIT Press.
8. Fagin, R., Halpern, J.Y. 1988. Belief, awareness, and limited reasoning, *Artificial Intelligence* 34, 39–76.
9. Gabbay, D., Johnson, R.H., Ohlbach, H.J., Woods, J. (eds), 2002. *Handbook of the Logic of Argument and Inference*, Amsterdam: North Holland.
10. Hintikka, J. 1962. *Knowledge and Belief*, Itchaca: Cornell University Press.
11. Hintikka, J. 1989. The Cartesian cogito, epistemic logic and neuroscience: some surprising interrelations, *Synthese* 83, 133–157.
12. Husserl, E. 1900–1/1970, *Logical Investigations*, London: Routledge.
13. Husserl, E. 1913/1982. *Ideas Pertaining to Pure Phenomenology and to a Phenomenological Philosophy* 1, Dordrecht: Kluwer.
14. Lakoff, G., Johnson, M. 1999. *Philosophy in the Flesh*, New York: Basic Books.
15. Manna, Z., Pnueli, A. 1993. Models for reactivity, *Acta Informatica* 30, 609–678.
16. Pietarinen, A.-V. 2002. Knowledge constructions for AI, Hacid, M.-S., et al., (eds), *Foundations of Intelligent Systems Intelligence*, Springer, 303–311.
17. Pietarinen, A.-V. 2003. What do epistemic logic and cognitive science have to do with each other?, *Cognitive Systems Research* 4, 169–190.
18. Pietarinen, A.-V. 2004a. Diagrammatic logic and game-playing, Malcolm, G. (ed.), *Multidisciplinary Studies of Visual Representations and Interpretations*, Elsevier.
19. Pietarinen, A.-V. 2004b. Semantic games in logic and epistemology, Gabbay, D. et al., (eds), *Logic, Epistemology and the Unity of Science*, Kluwer.
20. Pietarinen, A.-V. 2004c. Towards cognitive informatics: awareness in the brain, logic and cognitive neuroscience, Wang, Y. (ed.), *Cognitive Informatics: Exploring the Natural Intelligence*, World Scientific.
21. Pietarinen, A.-V. 2004d. Peirce's magic lantern of logic: moving pictures of thought, *Transactions of the Charles S. Peirce Society*.
22. Pietarinen, A.-V. 2004e. Multi-agent systems and game theory: a Peircean manifesto, *International Journal of General Systems*.
23. Pietarinen, A.-V. 2004f. The evolution of semantics and language games for meaning, to appear.
24. Rantala, V. 1982. Impossible worlds and logical omniscience, *Acta Philosophica Fennica* 35, 106–115.
25. Varela, F.J., Thompson, E., Rosch, E. 1991. *The Embodied Mind*, Cambridge, MA: MIT Press.
26. Wansing, H. 1990. A general possible worlds framework for reasoning about knowledge and belief, *Studia Logica* 49, 523–539.
27. Weiskrantz, L. 1997. *Consciousness Lost and Found: A Neuropsychological Exploration*, Oxford: Oxford University Press.
28. Young, A.W. 1994. Neuropsychology of awareness, in Revonsuo, A., Kamppinen, M. (eds), *Consciousness in Philosophy and Cognitive Neuroscience*, Hillsdale: Lawrence Erlbaum, 173–203.

Spinoza's Ontology

Jan Hladik*
hladik@tcs.inf.tu-dresden.de

Lehrstuhl für Automatentheorie, TU Dresden

Abstract. We examine the possibility of applying knowledge representation and automated reasoning in the context of philosophical ontology. For this purpose, we use the axioms and propositions in the first book of Spinoza's *Ethics* as knowledge base and a tableau-based satisfiability tester as reasoner. We are able to reconstruct most of Spinoza's system with formal logic, but this requires additional axioms which are assumed implicitly by Spinoza. This study illustrates how tools developed in computer science can be of practical use for philosophy.

1 Introduction

In this paper, we derive an ontology (knowledge base) from the ontology (metaphysics) of a philosopher. The problem of such an endeavour obviously lies in the fact that not only the word "ontology", but also most of the terminology related to reasoning is used differently in philosophy and knowledge representation. However, there are cases in which philosophers used the same formal strictness as in mathematics (or computer science): e.g., Baruch de Spinoza claims to have proven the propositions in his work *Ethics* [2] "in the geometric manner", i.e. with the same strictly logical approach as Euclid used in the *Elements*, his fundamental work on geometry and mathematics in general.

As expected, the translation of Spinoza's system into formal logic was not straightforward: in addition to the problems resulting from ambiguous formulation, it turned out that most of the theorems could not be proved with the literal translation of the axioms alone. In these cases, the examination of the proofs often revealed additional assumptions which Spinoza implicitly made and apparently considered as trivial. Since our intention was not to demonstrate the shortcomings of his work, but to reconstruct his ontology as adequately as feasible, we added these implicit axioms wherever possible to our knowledge base. Thus, the question examined in this paper is "How can the axioms be interpreted in such a way that the proofs hold" rather than "Which of the proofs hold with the intuitive interpretation of the axioms". We also do not attempt to reflect the historical discussion about the Ethics or the terms used (e.g. "substance") or discuss the legitimacy of the axioms. As we cannot examine all 36 propositions of the first book in detail, we restrict ourselves to the first 15 ones since they cover the essential statements of the first book [6] and climax in the principal statement that everything exists in and is conceivable through god.

* The author is supported by the DFG, Project No. GR 1324/3-4.

2 Translation of axioms and definitions

The structure of the Ethics is very clear: after sequences of definitions and axioms, a sequence of propositions together with proofs follows. Each of these is assigned a number, and thus we refer to a definition, axiom or proposition by the letters D, A, or P, respectively, together with its number.

Since our aim is to *effectively* decide the validity of Spinoza's proofs, we cannot use full first order predicate logic (FO), since it is undecidable. Instead, we use a decidable fragment of FO, namely the *Guarded Fragment* ($\mathcal{GF}$) [1, 4] which restricts the appearance of quantifiers to formulas of the kind

$$\forall \boldsymbol{x}(G(\boldsymbol{x},\boldsymbol{y}) \rightarrow \varphi(\boldsymbol{x},\boldsymbol{y})) \quad \text{or} \quad \exists \boldsymbol{x}(G(\boldsymbol{x},\boldsymbol{y}) \wedge \varphi(\boldsymbol{x},\boldsymbol{y})),$$

where $\boldsymbol{x}$ and $\boldsymbol{y}$ are tuples of variables; $G(\boldsymbol{x},\boldsymbol{y})$, called the *guard* of the formula, is an atom with variables $\boldsymbol{x}$ and $\boldsymbol{y}$; and $\varphi(\boldsymbol{x},\boldsymbol{y})$, called the *body* of the formula, is a guarded formula whose free variables are (a subset of) $\boldsymbol{x}$ and $\boldsymbol{y}$. Satisfiability of $\mathcal{GF}$ formulas is decidable in 2-ExpTime [3]. $\mathcal{GF}$ is quite expressive: for example, it allows for relations of arity greater than two, and it can express general axioms of the kind $\forall x\, \varphi(x)$ for a formula φ with one free variable x in the following way: $\forall x((x = x) \rightarrow \varphi(x))$.

For $\mathcal{GF}$, there exists the tableau-based satisfiability tester Saga [5]. A satisfiability tester can be used to check the validity of an implication $\varphi \Rightarrow \psi$ in the following way: if the formula $\varphi \wedge \neg\psi$ is satisfiable, then the implication is not valid. Through this procedure, no formal proof is obtained for a valid implication, but in the opposite case a counter-model is found, which can be used to examine the reason for the failure of the implication. Moreover, since this algorithm is a decision procedure, we can be sure that if no model is found by the algorithm, then there exists none and therefore the implication is valid. One of the main practical advantages of tableau algorithms is their performance, which is much better in practice than their worst-case complexity suggests.

Although Spinoza's style bears a high similarity with mathematical language, his statements require interpretation before they can be translated into formal logic. We will illustrate this with two examples. Firstly, there is no clear distinction between universal and existential quantification: by D3, "a substance is conceivable through itself". If this is interpreted as existential quantification, it reads: "Substance is *at least* conceivable through itself (but possibly also through something different)." With universal quantification, it means: "Substance is *only* conceivable through itself (but possibly not conceivable at all)." In this case and in several similar ones, we think that only both statements together reflect Spinoza's intention, since this eliminates the possibilities in the parentheses. Secondly, in the Latin text there are two words which translate into English as "or": "vel" (e.g. A1, P4) stands for a logical alternative, and therefore has to be translated into formal logic as "$\vee$". In contrast, "sive" (e.g. D1, D4, A5) means rather "or also", i.e. it is followed by an alternative definition. Thus logically, this "or" has to be translated into "$\wedge$".

In order to express the axioms and definitions, we use nine unary, eight binary and two ternary relations, which are explained in Table 1. In the following, we present our translation of the definitions and axioms. Starting with (the English translation of) Spinoza's original formulation, we present our understanding first in ordinary language, then in formal logic.

Table 1. Relations

Unary Relations	Meaning	Bi-/ternary Relations	Meaning
$\mathsf{S}(x)$	x is a substance	$\mathsf{sn}(x,y)$	x has the same nature as y
$\mathsf{A}(x)$	x is an attribute	$\mathsf{ao}(x,y)$	x is an attribute of y
$\mathsf{M}(x)$	x is a mode	$\mathsf{mo}(x,y)$	x is a mode of y
$\mathsf{E}(x)$	x is existing	$\mathsf{ei}(x,y)$	x exists in y
$\mathsf{C}(x)$	x is conceivable	$\mathsf{ct}(x,y)$	x is conceivable through y
$\mathsf{G}(x)$	x is (a) god	$\mathsf{sc}(x,y)$	x has something in common with y
$\mathsf{F}(x)$	x is finite in its kind	$\mathsf{lb}(x,y)$	x can be limited by y
$\mathsf{CA}(x)$	x is a cause	$\mathsf{co}(x,y)$	x is cause of y
$\mathsf{EF}(x)$	x is an effect	$\mathsf{ca}(x,y,z)$	x and y have the common attribute z
		$\mathsf{di}(x,y,z)$	x can be divided into y and z

D1 *By that which is "self-caused" I mean that of which the essence involves existence, or that of which the nature is only conceivable as existent.* This is rather an axiom than a definition, since "self-caused" is a complex term: something has itself and only itself as cause if and only if its existence follows from its conceivability. Note that this implies that everything which exists and everything which is not conceivable is self-caused.

$$\forall x((x=x)\rightarrow((\mathsf{co}(x,x)\wedge(\forall y(\mathsf{co}(y,x)\rightarrow(x=y))))\leftrightarrow(\mathsf{C}(x)\rightarrow(\mathsf{E}(x)))))$$

D2 *A thing is called "finite in*[1] *its kind" when it can be limited by another thing of the same nature [...]* A thing is finite in its kind if and only if there exists a different thing with the same nature by which it can be limited.

$$\forall x((x=x)\rightarrow(\mathsf{F}(x)\leftrightarrow\exists y(\mathsf{lb}(x,y)\wedge\mathsf{sn}(x,y)\wedge(x\neq y))))$$

D3 *By "substance" I mean that which is in itself, and is conceived through itself [...]* A substance exists in itself (and only in itself), and it is conceivable through itself (and only through itself).

$$\forall x((x=x)\rightarrow(\mathsf{S}(x)\leftrightarrow\mathsf{ei}(x,x)\wedge\mathsf{ct}(x,x)\wedge\forall y(\mathsf{ei}(x,y)\rightarrow(x=y))\wedge\forall y(\mathsf{ct}(x,y)\rightarrow(x=y))))$$

D4 *By "attribute" I mean that which the intellect perceives as constituting the essence of substance.* An attribute is conceivable, and it is an attribute of a substance (and only of substances), which are also conceivable. Moreover, every substance has an attribute.

$$\forall x(\mathsf{S}(x)\rightarrow\exists y(\mathsf{ao}(y,x)))$$

$$\forall x((x=x)\rightarrow(\mathsf{A}(x)\leftrightarrow\mathsf{C}(x)\wedge\exists y(\mathsf{ao}(x,y))\wedge\forall y(\mathsf{ao}(x,y)\rightarrow(\mathsf{S}(y)\wedge\mathsf{C}(y)))))$$

D5 *By "mode" I mean the modifications ("affectiones") of substance, or that which exists in, and is conceived through, something other than itself.* A mode exists (only) in and is (only) conceivable through a substance.

$$\forall x((x=x)\rightarrow(\mathsf{M}(x)\leftrightarrow\mathsf{C}(x)\wedge\exists y(\mathsf{mo}(x,y))\wedge\forall y(\mathsf{mo}(x,y)\rightarrow(\mathsf{ei}(x,y)\wedge\mathsf{ct}(x,y)\wedge\mathsf{S}(y)))))$$

D6 *By "God" I mean a being absolutely infinite—that is, a substance consisting in infinite attributes, of which each expresses eternal and infinite essentiality.* If a god is absolutely infinite, he must also be infinite in his kind. We omit the predicate "eternal", since it is not used in any other proposition considered in this paper. Regarding the meaning of "infinite attributes", see our explanation for I1 below.

$$\forall x((x=x)\rightarrow(\mathsf{G}(x)\leftrightarrow(\mathsf{S}(x)\wedge\neg\mathsf{F}(x))))$$

[1] We think that the literal translation "in" reflects Spinoza's intention better than "after".

A1 *Everything which exists, exists either in itself or in something else.* Everything existing exists in something, which leaves open the possibility that it is something else or the same. $\forall x(\mathsf{E}(x) \rightarrow \exists y(\mathsf{ei}(x,y)))$

A2 *That which cannot be conceived through anything else must be conceived through itself.* Everything conceivable is conceivable through something (possibly itself) which, as is suggested by D3 and D5, must be a substance.
$\forall x(\mathsf{C}(x) \rightarrow \exists y(\mathsf{ct}(x,y) \wedge \mathsf{S}(y)))$

A3 *From a given definite cause an effect necessarily follows; and, on the other hand, if no definite cause be granted, it is impossible that an effect can follow.* Every cause has an effect, and vice versa.
$\forall x(\mathsf{CA}(x) \rightarrow \exists y(\mathsf{co}(x,y) \wedge \mathsf{EF}(y)))$ $\forall x(\mathsf{EF}(x) \rightarrow \exists y(\mathsf{co}(y,x) \wedge \mathsf{CA}(y)))$

A4 *The knowledge of an effect depends on and involves the knowledge of a cause.* Since "knowledge" does not appear in any other axiom or definition and is used synonymously with "conception" in the corollary of P6, we treat the two terms as synonymous. Thus, something is a conceivable effect if and only if it has a cause and is conceivable through its causes.
$\forall x((x = x) \rightarrow ((\mathsf{EF}(x) \wedge \mathsf{C}(x)) \leftrightarrow (\exists y(\mathsf{co}(y,x)) \wedge \forall y(\mathsf{co}(y,x) \rightarrow (\mathsf{ct}(x,y))))))$

A5 *Things which have nothing in common cannot be understood, the one by means of the other; the conception of one does not involve the conception of the other.* Expressed positively, this means: if x is conceivable through y, then x and y have something in common. $\forall x,y(\mathsf{ct}(x,y) \rightarrow \mathsf{sc}(x,y))$

We omit D7, D8 and A6 since they are not used for the propositions considered in this paper, and A7 because it is merely a repetition of D1.

As mentioned in Section 1, we added implicit axioms, i.e. axioms which follow immediately from the semantics of the words used (e.g. if x is the cause of y, then x is a cause and y is an effect), or assumptions which are only revealed in the proofs (e.g. in the proof of P4, it is argued that substances and modes are the only existing things). In the following list, we enumerate and justify these additional axioms.

I1 In D6, god is described as a substance having "infinite attributes". In the proof of P14, it becomes clear that not the attributes themselves are infinite, but their number, or more precisely that god has all attributes: "god is a being [...], of whom no attribute [...] can be denied (by D6)". Thus, every attribute is an attribute of god. Moreover, since every substance has an attribute, it has a common attribute with god.[2]
$\forall x(\mathsf{A}(x) \rightarrow \exists y(\mathsf{ao}(x,y) \wedge \mathsf{G}(y)))$ $\forall x(\mathsf{S}(x) \rightarrow \exists y,z(\mathsf{ca}(x,y,z) \wedge \mathsf{G}(y)))$

I2 In the proof of P4, Spinoza states that "outside[3] of the intellect, there is nothing except substance and its modifications." Hence, everything existing is either substance or mode. $\forall x(\mathsf{E}(x) \rightarrow (\mathsf{S}(x) \vee \mathsf{M}(x)))$

I3 In the proof of P2, Spinoza (implicitly) assumes that if two things are not conceivable through each other, then they have nothing in common, i.e. that the implication of A5 also holds in the opposite direction. Therefore, we also add

[2] We have to state this explicitly since we cannot fully express P5 in $\mathcal{GF}$, see our explanation for P5 below.

[3] We think that translating "extra" literally as "outside" instead of "in addition to" makes more sense in this context, since then "extra intellectum" simply means "in nature", whereas otherwise its meaning becomes unclear (see [7], p. 14).

the axiom that if two things have something in common, then one is conceivable through the other. $\forall x, y(\mathsf{sc}(x,y) \rightarrow (\mathsf{ct}(x,y) \vee \mathsf{ct}(y,x)))$

I4 In P5, Spinoza mentions "substances having the same nature or ('sive') attribute", which indicates that the term "same nature" in D2 has to be interpreted as "same attribute". For this purpose, we introduce the relation $\mathsf{ca}(x,y,z)$, with the semantics that both x and y have attribute z, and they have something in common (namely z). $\forall x, y(\mathsf{sn}(x,y) \rightarrow \exists z(\mathsf{ca}(x,y,z)))$
$\forall x, y, z(\mathsf{ca}(x,y,z) \rightarrow (\mathsf{ao}(z,x) \wedge \mathsf{ao}(z,y) \wedge \mathsf{co}(x,y)))$

I5 The relation sc is symmetric and reflexive.
$\forall x, y(\mathsf{sc}(x,y) \rightarrow \mathsf{sc}(y,x))$ $\forall x((x = x) \rightarrow \mathsf{sc}(x,x))$

I6 If x is conceivable through (or exists in) y, then both x and y are conceivable (or existing, respectively). If x is the cause of y, then x is a cause and y is an effect. Analogous axioms hold for the relations ao and mo.
$\forall(x,y)(\mathsf{ct}(x,y) \rightarrow (\mathsf{C}(x) \wedge \mathsf{C}(y)))$ $\forall(x,y)(\mathsf{ao}(x,y) \rightarrow (\mathsf{A}(x) \wedge \mathsf{S}(y)))$
$\forall(x,y)(\mathsf{ei}(x,y) \rightarrow (\mathsf{E}(x) \wedge \mathsf{E}(y)))$ $\forall(x,y)(\mathsf{mo}(x,y) \rightarrow (\mathsf{M}(x) \wedge \mathsf{S}(y)))$
$\forall(x,y)(\mathsf{co}(x,y) \rightarrow (\mathsf{CA}(x) \wedge \mathsf{EF}(y)))$

I7 The sets of substances, attributes and modes are pairwise disjoint.
$\forall x(\mathsf{S}(x) \rightarrow \neg\mathsf{A}(x))$ $\forall x(\mathsf{S}(x) \rightarrow \neg\mathsf{M}(x))$ $\forall x(\mathsf{A}(x) \rightarrow \neg\mathsf{M}(x))$

I8 We express the relation "x is divisible into y and z" as used in P12 and P13 as "x is cause of two different things y and z, but not equal to either."
$\forall x, y, z(\mathsf{di}(x,y,z) \rightarrow (\mathsf{co}(x,y) \wedge \mathsf{co}(x,z) \wedge (x \neq y) \wedge (x \neq z) \wedge (y \neq z)))$

3 Translation and proof of propositions

In this section, we consider the propositions in detail. As in Section 2, we first give the original formulation by Spinoza, followed by our interpretation and the translation into $\mathcal{GF}$. In order to improve readability, we use the notation $\alpha \rightarrow \beta$ for $\neg\alpha \vee \beta$ (material implication), and $\varphi \Rightarrow \psi$ for the (logical) implication claimed in the proposition, though the translation of both into $\mathcal{GF}$ is identical. Consequently, the satisfiability of $\varphi \wedge \neg\psi$ was tested by SAGA.

P1 *Substance is by nature prior to its modifications.* The term "prior" is not formally defined in the axioms and definitions. If we interpret it logically and not temporally (which is suggested by the proof, since time is not mentioned in D3 and D5), we obtain the following: if there exist a mode, then there also exists a substance. This proposition really follows from D5, and thus SAGA cannot find a counter-model. $\mathsf{M}(x) \Rightarrow \exists y(\mathsf{S}(y))$

P2 *Two substances, whose attributes are different, have nothing in common.* If z is an attribute of the substance x, but not of the substance y, then x and y have nothing in common. Here, Spinoza only proves that the conception of x does not involve the conception of y. Without I3, SAGA finds a counter-model with two substances which have something in common, but are not conceivable through each other. If we add I3, this proposition follows from the fact that in order to have something in common, one of the substances has to be conceivable through the other one, but every substance is only conceivable through itself.
$(\mathsf{S}(x) \wedge \mathsf{S}(y) \wedge \mathsf{A}(z) \wedge \mathsf{ao}(z,x) \wedge \neg\mathsf{ao}(z,y)) \Rightarrow \neg\mathsf{sc}(x,y)$

P3 *Things which have nothing in common cannot be one the cause of the other.* Spinoza's proof relies on A4, but this axiom is only applicable if conceivability of the effect is assumed. However, if the effect is not conceivable, it is (by

D1) self-caused, and thus (by I5) has something in common with its cause, so the proposition also holds in this case. $\neg\mathsf{sc}(x,y) \Rightarrow \neg\mathsf{co}(x,y)$

P4 *Two or more distinct things are distinguished one from the other, either by the difference of the attributes of the substances, or by the difference of their modifications.* Two things x, y having a common attribute z and a common mode w are identical. This proposition holds; the explanation is similar to the one for P5 below. $(\mathsf{ca}(x,y,z) \wedge \mathsf{mo}(w,x) \wedge \mathsf{mo}(w,y)) \Rightarrow (x = y)$

P5 *There cannot exist in the universe two or more substances having the same nature or attribute.* The intuitive translation of this proposition into FO is: $\forall x, y, z((\mathsf{S}(x) \wedge \mathsf{S}(y) \wedge \mathsf{A}(z) \wedge \mathsf{ao}(z,x) \wedge \mathsf{ao}(z,y)) \Rightarrow (x = y))$, which is not guarded, since x and y appear together in the body, but not in the guard. In order to express this in $\mathcal{GF}$, we introduce the relation ca and the axiom I4. Then, we can express the proposition as follows: two substances with a common attribute are identical. This results from P2 and therefore also depends on I3: without this additional axiom, SAGA finds a counter-model with two finite substances having a common attribute. $(\mathsf{S}(x) \wedge \mathsf{S}(y) \wedge \mathsf{ca}(x,y,z)) \Rightarrow (x = y)$

P6 *One substance cannot be produced by another substance.* The proof only shows that one substance cannot be the cause of another one, which suggests that the terms "cause of" and "produced by" are used synonymously, and thus we can express this as follows: if two substances are different, then one cannot be the cause of the other; this follows from A4. $(\mathsf{S}(x) \wedge \mathsf{S}(y) \wedge (x \neq y)) \Rightarrow \neg\mathsf{co}(x,y)$

P7 *Existence belongs to the nature of substances.* Every substance is existing. This follows directly from D3 and I6. $\mathsf{S}(x) \Rightarrow \mathsf{E}(x)$

P8 *Every substance is necessarily infinite.* This indeed follows from D2 and P5, which state that a thing can only be limited by another thing of the same nature, but substances with the same nature are identical. $\mathsf{S}(x) \Rightarrow \neg\mathsf{F}(x)$

P9 *The more reality or being a thing has, the greater the number of its attributes.* We do not see a possibility of expressing this in $\mathcal{GF}$ or even in FO. Since this proposition is not used in any other proof in the first book of the ethics, it probably serves merely as a justification for a substance to have several attributes. Hence, we do not consider it any further.

P10 *Each particular attribute of the one substance must be conceived through itself.* Spinoza's proof shows only that the attribute of a substance must be conceived through this substance. Still, no counter-model can be found: by I5, everything has something in common with itself, and by I3, this implies that it is conceivable through itself. $\mathsf{A}(x) \Rightarrow \mathsf{ct}(x,x)$

P11 *God [...] necessarily exists.* There can be no model in which there is no individual with the predicate G. To avoid an empty structure[4], we add an individual x with the predicate $C(x)$, so that our proposition reads "If something is conceivable, then there exists a god", which also reflects the structure of the ontological argument used by Spinoza.[5] $C(x) \Rightarrow \exists y(\mathsf{G}(y))$

P12 *No attribute of substance can be conceived from which it would follow that substance can be divided.* The main problem of this and the next proposition is the interpretation of "division". The reference to P6 indicates that if a

[4] In an empty structure, even a formula like $\forall x(C(x) \wedge \neg C(x))$ is satisfiable.

[5] Note that the naive translation $\mathsf{G}(x) \Rightarrow \mathsf{E}(x)$ follows directly from the definition of substance, and it does not show that god *necessarily* exists.

substance x is divided into y and z, then it must be the cause of y and z, but not equal to either. Then, neither y nor z can be a substance (by P6), nor can both be something different and thus x cease to exist (P7).

$$\mathsf{A}(w) \land \mathsf{C}(w) \land \mathsf{S}(x) \land \mathsf{ao}(w,x) \land \mathsf{di}(x,y,z) \Rightarrow \neg\mathsf{S}(y) \land \neg\mathsf{S}(z) \land \mathsf{E}(x)$$

P13 *Substance absolutely infinite is indivisible.* The absolutely infinite substance is god. Beyond this, P13 is merely a repetition of P12 without the attribute w. $\mathsf{G}(x) \land \mathsf{di}(x,y,z) \Rightarrow \neg\mathsf{S}(y) \land \neg\mathsf{S}(z) \land \mathsf{E}(x)$

P14 *Besides God no substance can be granted or conceived.* The argument relies on the assumptions that god has all attributes (D6, I1), that two substances x and y which have something (e.g. an attribute) in common are conceivable through each other (I3), and that every substance is only conceivable through itself (D3), such that x and y must be identical. Again, without I3 there exists a counter-model with a finite substance that has a common attribute with god, but is not conceivable through him. $\mathsf{S}(x) \land (\mathsf{E}(x) \lor \mathsf{C}(x)) \Rightarrow \mathsf{G}(x)$

P15 *Whatsoever is, is in God, and without God nothing can be, or be conceived.* This follows quite easily from P14, since everything existing (i.e. substances and modes) exists in and is conceivable through a substance, and god is the only substance. $(\mathsf{E}(x) \Rightarrow \exists y(\mathsf{ei}(x,y) \land \mathsf{G}(y))) \land (\mathsf{C}(x) \Rightarrow \exists y(\mathsf{ct}(x,y) \land \mathsf{G}(y)))$

Summary. Out of the 15 propositions under consideration, 12 can be proved using Spinoza's axioms and definitions and the implicit axioms mentioned above. P11 can be proved with some appropriate pre-condition. P5 cannot be expressed fully in $\mathcal{GF}$, but it can be reformulated in a way that renders it provable. Only one proposition remains: the meaning of P9 is very vague and cannot be expressed in $\mathcal{GF}$. The implicit axiom I3 ("if two things have something in common, then one of them is conceivable through the other") is crucial for the high number of provable propositions. Although it is not explicitly formulated by Spinoza and its legitimacy is questionable from a philosophical point of view, it is clearly assumed by Spinoza in the proof of P2, and without it, counter-models for the propositions 2, 4, 5, 8, 10, 14 and 15 can be found. This dependency is also reflected in Spinoza's proofs: P14 relies on P5, which in turn relies on P2.

In Figure 1, the subsumption hierarchy of the unary relations is shown: an arrow $A \rightarrow B$ indicates that the relation A is *subsumed* by B, i.e. that every instance of A is an instance of B. To improve readability, we show only direct subsumption; note that the subsumption relation is transitive and thus $A \rightarrow B \rightarrow C$ implies $A \rightarrow C$. Here, the symbol $\bot$ stands for an inconsistent relation, which is subsumed by all other ones, and $\top$ is the universal relation, which subsumes all other ones. The trivial arrows from $\bot$ and to $\top$ are shown as dashed. If two relations mutually subsume each other, they are displayed as one.

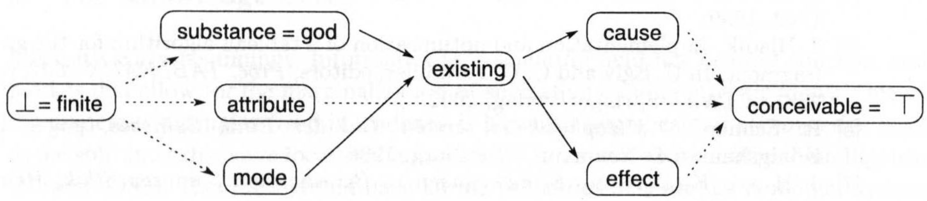


Fig. 1. Subsumption hierarchy

This hierarchy reveals additional propositions which follow from the set of axioms but are not mentioned or proved by Spinoza. Firstly, everything is conceivable, since everything has something in common with itself (I5), and is therefore conceivable through itself (I3). This result again demonstrates the problematic implications of I3. Secondly, 'finite' is inconsistent. It has been shown that substances are infinite (P8), but why cannot something else, e.g. a mode, be finite? The reason for this is D2, which defines 'finite' as something which can be limited by some other thing of the same nature. By I4, 'nature' means 'attribute', but only substances have attributes. Thus, anything other than a substance cannot be finite either. Thirdly, everything existing is cause as well as effect. This results from D1, since everything existing is self-caused. All other subsumptions are mentioned more or less explicitly by Spinoza. This is an indicator that our system is not over-constrained, i.e. through our translation and the additional axioms we did not introduce any conclusions which do not follow from Spinoza's system.

4 Conclusion

We developed an ontology based on the arguments made in the first part of Spinoza's *Ethics*. While testing the validity of the proofs provided, it turned out that most of them require additional axioms, which are only revealed inside the proofs. In particular, the main theorem depends crucially on an implicit assumption which is not justified anywhere. The counter-models provided by SAGA proved to be very helpful in the detection of these missing prerequisites. With the additional axioms, we are able to prove most of the propositions. The restrictions imposed by the language $\mathcal{GF}$ cause problems only in one case, and a workaround is possible. In addition to the propositions claimed by Spinoza, we are able to derive additional theorems following from the axioms. In summary, we think that this illustrates the use of automated reasoning tools for philosophy.

References

[1] H. Andréka, J. van Benthem, and I. Németi. Modal languages and bounded fragments of predicate logic. *J. of Philosophical Logic*, 27(3):217–274, 1998.

[2] B. de Spinoza. *Ethics*. Dover Publications, New York, 1951. Online available from `http://www.gutenberg.net/browse/BIBREC/BR004571.HTM`, a bilingual version can be found on `http://home.tiscali.be/rwmeijer/spinoza`.

[3] E. Grädel. Decision procedures for guarded logics. In H. Ganzinger, editor, *Proc. of CADE-16*, volume 1632 of *LNAI*. Springer-Verlag, 1999.

[4] E. Grädel. On the restraining power of guards. *Journal of Symbolic Logic*, 64:1719–1742, 1999.

[5] J. Hladik. Implementation and optimisation of a tableau algorithm for the guarded fragment. In U. Egly and C. G. Fermüller, editors, *Proc. TABLEAUX 2002*, volume 2381 of *LNAI*. Springer-Verlag, 2002.

[6] R. Schnepf. *Metaphysik im ersten Teil der Ethik Spinozas*, pages 17–18. Königshausen & Neumann, Würzburg, 1996.

[7] J. H. von Kirchmann. *Erläuterungen zu Benedict von Spinozas Ethik.* Heimann, Berlin, 1871.

Knowledge Relativity

Oliver Hoffmann oliver@hoffmann.org

University of South Australia

Abstract. I introduce the notion of *knowledge relativity* as a proposed conceptual link between different scientific disciplines. Examples from Informatics and Philosophy, particularly Newell's *knowledge level* hypothesis and Popper's *world 3* of *knowledge*, are used to demonstrate the motivation for making this notion explicit.

1 Why Knowledge Relativity?

Not every assumption is *knowledge* and whether somebody's personal views on the world are classified as *knowledge* is determined by complicated interactions among human subjects and between human subjects and their environment.

The word *science* itself means *knowledge* and thus one would assume that the distinction between arbitrary unilateral assumptions on the one hand and shared mutually assured *knowledge* on the other hand is uncontested in contemporary science. But one scientific discipline has adopted a notion of *knowledge* that abandons the distinction between *knowledge* itself on the one hand and the mode of it's expression or temporary attempts to arrive at it on the other hand: In Informatics, *knowledge representation* is progressively equated with *knowledge* itself, and isolated pieces of *knowledge representation* stored in a single computer system with mutually confirmed *truth*. Philosophy would be ideally situated to contribute to a more mature notion of *knowledge* in Informatics, but the relativity of *knowledge* appears to be too commonplace in contemporary Philosophy to be explicitly stated.

In this article, I will therefore try to explicate *knowledge relativity* and the potential role of this notion both in Informatics and in Philosophy. The starting point for this task will be asking the missing "Who" question for some of the basic concepts in Informatics and the result will be a notion of *knowledge* that emphasizes *subjects* over *objects* and involves an understanding of *standardization* instead of the predominant notion of *truth*.

2 The Knowledge Level

Since it's early beginnings, Informatics as a scientific field has favored concepts and models that allow for the marginalization of subjectivity and relativistic views. One of the numerous examples for this tendency is Newell's hypothesis of a *knowledge level* as the sole and exhaustive location for *knowledge* in a system with artificial intelligence (Newell 1982). This level would have all the properties of a normal computer system level: among other properties, it can be implemented without references to internal details of other levels, and it can be reduced to the level below it (the *symbol level*) by defining it's medium (*knowledge*), components and laws via those (*symbols*) of the level below it. According to this hypothesis, *knowledge* could be seen as an abstract property of (computer system or human) *agents* implementable via various different forms of symbolic representations in the same way that symbolic representations are implementable via various different forms of electronic hardware. Such an understanding of *knowledge* would make it impossible to directly verify or falsify the presence of a specific element of *knowledge* in an artificial *agent*: the highest directly observable system level contains symbols that might or might not encode a specific *knowledge*, but the *knowledge* itself would reside one level above those representations and would be removed from direct observation. But Newell proposed a mechanism of indirect verification for the *knowledge level*: If some (human or artificial) *agent* A can detect the impact of some specific *knowledge* in the actions of another *agent* B, *agent* A can verify the presence of such *knowledge* in *agent* B (Fig. 1). Through this hypothesis, Newell

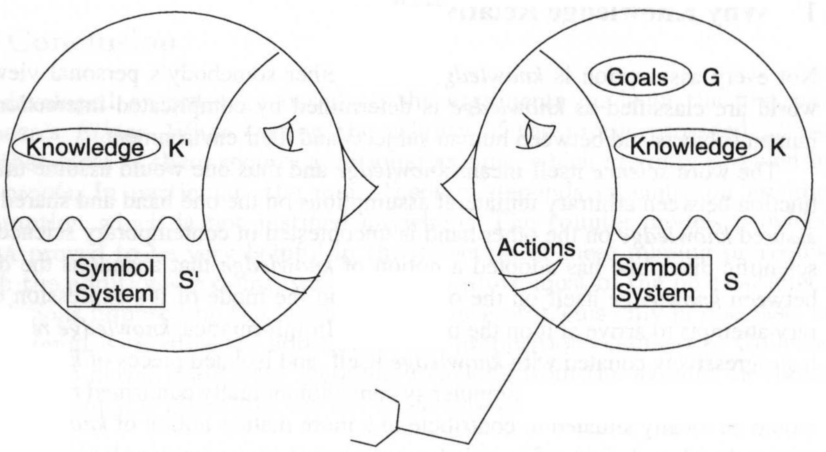


Fig. 1. verification of knowledge according to Newell

attempted to end an internal dispute among artificial intelligence researchers, a controversy centered around the question of the "best" form of *knowledge representation*. According to Newell himself (Newell 1993), this attempt was partially successful, and a significant part of artificial intelligence research has been (implicitly or explicitly) based on the *knowledge level hypothesis*. I will not discuss the correctness of Newell's hypothesis in this article, partially because it is too complex to determine how it could be verified or falsified at all, but also because the correctness or incorrectness of this hypothesis is not directly linked to the central argument in this article: the general tendency of Informatics to suppress notions of relativity and the need to introduce explicit terminology for expressing relativity. A starting point for detecting the implicit *objectivist* world view (Stamper 1993) underlying artificial intelligence concepts like the *knowledge level* is to focus on those aspects that authors like Newell do not discuss,

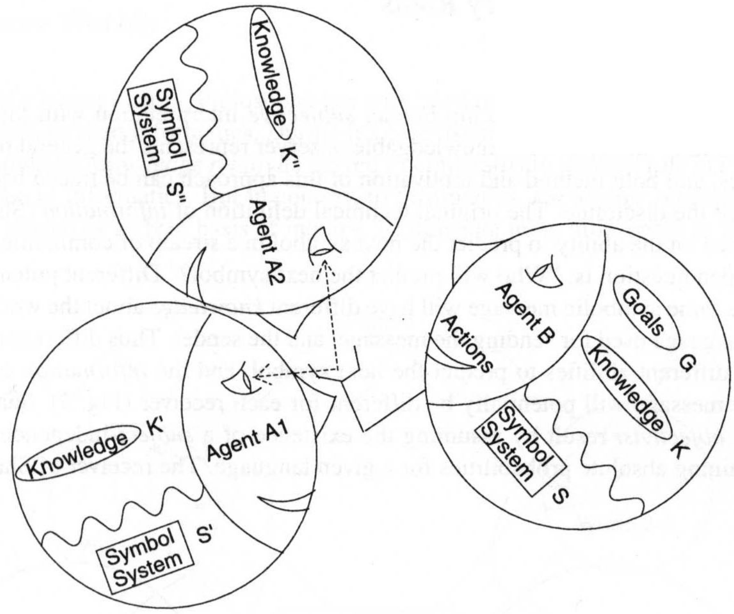


Fig. 2. how do different observing agents agree on the presence of knowledge?

and the questions that are not answered. A good example is: "Who is the *agent* that will verify the presence of *knowledge* in another *agent*?" More than one *agent* could take the role of the observing agent A in Newell's *knowledge* verification mechanism and the different As could come to different conclusions on whether *knowledge* is present in *agent* B (Fig. 2). Newell does not explain how the A *agents* would be able to come to any consistent conclusion on *agent* B's knowledge, and since their symbol levels might be mutually incompatible, it is unclear how they could even communicate their different views on *agent* B. The only conceivable way Newell's *knowledge* verification might yield consistent results is by relying on a standardized observer *agent* A that serves as the absolute reference on detecting *knowledge*. Newell implicitly assumes *objective knowledge* that would be the basis, rather than the result of attempts to create artificial intelligence. But if such a standard *agent* existed, Newell's entire *knowledge* verification procedure might be obsolete before it is ever applied: this standard verification *agent* would already incorporate the perfect embodiment of intelligence and *knowledge*, and any attempts to build more intelligent *agents* would have to fail. The main effect of Newell's *knowledge level hypothesis* is not what it accomplishes, but what it prevents: the authority to interpret symbolic *knowledge representations* is restricted to computer systems, since their (virtual) behavior is the only way for determining what *knowledge* is represented. Human subjective interpretation is excluded from the process of symbol interpretation, since Newell's artificial *agents* never directly expose their symbol level to human *agents*.

3 Information Theory Roots

Newell's approach of replacing human *subjective* interpretation with the implicit assumption of an *objectively* knowledgable observer represents the general trend in Informatics, and both method and motivation of this approach can be traced back to the origins of the discipline: The original technical definition of *information* (Shannon 1948) is based on the ability to predict the next symbol in a stream of communication. Again, the open question is: "Who will predict the next symbol?" Different potential receivers of the same symbolic message will have different *knowledge* about the world in general, the language used for sending the message, and the sender. Thus different receivers will have different abilities to predict the next symbol, and the *information* content of the same message will potentially be different for each receiver (Fig. 3). Shannon arrives at an *objectivist* result by assuming the existence of a *subject*-independent dictionary containing absolute probabilities for a given language. The receiver of Shannon's messages turns out to be the same implicitly standardized perfectly knowledgable observer that Newell uses to verify the existence of *knowledge* in computer systems. Shannon, like Newell, was motivated by the goal of eliminating subjective interpretation from his model of symbolic message content, and his *information* notion is still in use today. Informatics has suppressed *knowledge relativity* since it's beginning, and current trends to equate *knowledge* with *knowledge representation* are the direct continuation of this tradition. The discipline initially had good reasons for choosing this approach: At the time of Shannon, technical reliability of communication was the primary goal, and side effects on the potential processing of stored *knowledge* inside computer systems were largely irrelevant to the engineers that founded Informatics. Wherever issues of *knowledge relativity* come in conflict with engineering properties like reliability, predictability, and consistency, Informatics has generally given preference to those views that favor engineering goals. Implicitly, *knowledge relativity* has always played a role in Informatics: as the notion that has to be eliminated.

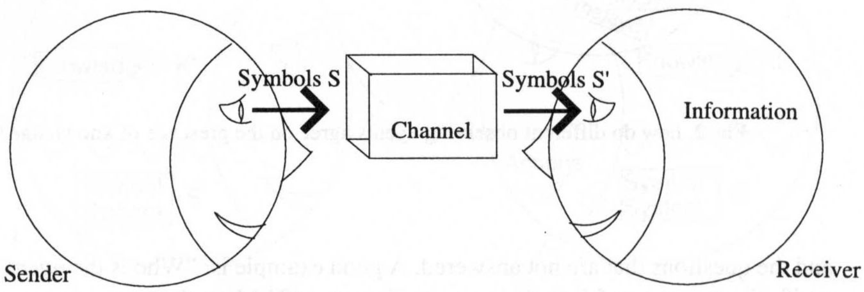


Fig. 3. different receiver *agents* would receive different amounts of Shannon *information*

4 Three Worlds

Due to it's longer history, Philosophy has witnessed more changes in the role of *knowledge relativity* than Informatics. During it's co-existence with Informatics in the last 60 years, however, *knowledge relativity* played a comparatively implicit role in Philosophy, much like in Informatics. But in contrast to Informatics, *knowledge relativity* was implicitly treated as a given basis of inquiry. To illustrate main differences in the implicit

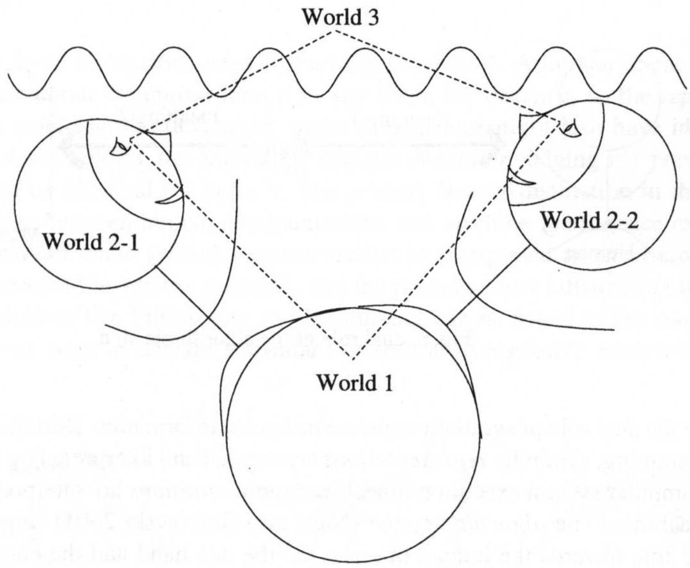


Fig. 4. Popper's *3 worlds* thesis

treatment of *knowledge relativity*, I will discuss a prominent example from the Philosophy of Science: Popper's *3 worlds* thesis (Popper 1972). Popper defines 3 different *worlds* that are interrelated but separate (Fig. 4): *World 1* is the (observer-independent) physical universe, *world 2* is an observer's image of *world 1*, and *world 3* is the collective symbolic *representation* of *world 2* shared among a multitude of observers. *World 3* is necessarily embedded in *world 1*, since observers are only able to communicate each other's symbolic *representations* if they are able to physically perceive them. Furthermore, any action performed by one of the observers can be interpreted as a contribution to *world 3*, and scientific experiments are the most prominent example of this feature. Popper's motivation for his *3 worlds* thesis was somewhat similar to Newell's motivation for his *knowledge level* hypothesis: Popper wanted to show a path towards *objective* observer-independent *knowledge*. But in contrast to predominant Informatics approaches, Popper assumed *subject*-relative *knowledge* and *subject*-relative *representation* as the starting point of this path. *Objective knowledge* is only achieved as the result of a long process of negotiation among *subjects*. This process is permanent, since *world 3* only approximates *world 1*, without ever reaching total consistency with it.

5 A Matter of Perspective

The reception of Philosophical concepts in Informatics is somewhat ambiguous: Informatics approaches frequently borrow their terminology - for instance *ontology* (Gruber 1991) - from philosophy, but later transform the meaning of these terms in very drastic ways, at times even reversing their original meaning. I argue that this phenomenon is

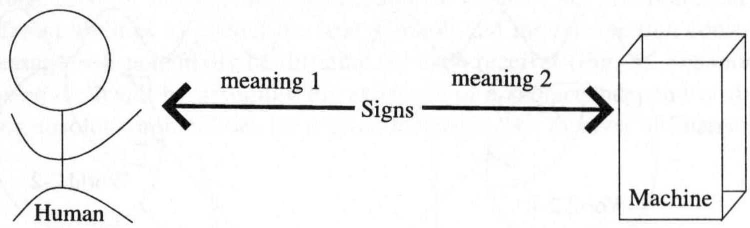


Fig. 5. dual role of the algorithmic sign

due to the dual role of symbolic *representations* in Informatics: During computer system programming, symbolic *representations* are created and interpreted by humans, but during computer system execution time, these *representations* are interpreted and modified by machines. The *algorithmic sign* (Nake and Grabowski 2001) (Fig. 5) incorporates a dual role towards the human observer on the one hand and the computer system on the other hand, at times as an external symbol used by human *agent*(s), and at times as an internal symbol used by machine(s). Machines are standardized products, and they are expected to work according to specifications. The same symbolic *representation* would therefore be expected to be open for re-interpretation when read and analyzed by different human *agents*, but would be expected to be stable and rigid in it's meaning and function when processed by different computer system *agents*. Since the final goal of Informatics is creating computer systems, the discipline has typically favored views that marginalize human *subject*-relative aspects. This approach has proven successful in areas that require predictability and consistency, but unsuccessful when the cost of implicit standardization was too high. One example for the latter are *knowledge* management tasks or the computer support of human creativity. For a more successful interaction with computer systems in these areas, *knowledge relativity* has to be made explicit.

6 Dimensions of Knowledge Relativity

In order to explicate the relativity of *knowledge* I will try to list the different dimensions of *knowledge relativity*:

1. *subject* who is holding the knowledge: who knows?
2. *representation* what kind of symbolic expression is used to communicate the knowledge: what was expressed?
3. *evaluator* who is judging the presence of knowledge: who thinks someone knows?
4. *communicative intent* was the *knowledge* expressed in order to inform or in order to reach agreement on known issues: do we need to discuss this?
5. *functional intent* what use was intended for the representation used: what will it change?
6. *receiver* who was the intended receiver for the *representation* used: who should know?

Not all of these dimensions need to be fully developed. A human agent might have *knowledge* without communicating it in any form, for instance, so the *representation* dimension would not be developed. Some dimensions might also have identical values, the *subject* holding the *knowledge* and the *evaluator* judging it's presence might for instance be identical ("I know"). The primary line of distinction in the treatment of *knowledge* between human communication and machine intelligence can be found in the *functional intent* (will the *representation* be interpreted as an external signal or directly processed inside the system?), and the primary line of distinction between different disciplines like Philosophy and Informatics can be found in the *communicative intent* (do we want to declare a standard or start a discussion?). Such a definition of

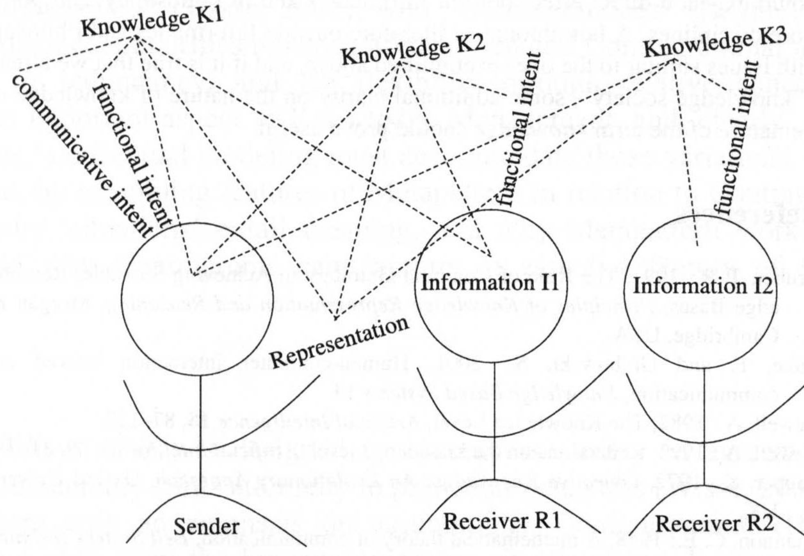


Fig. 6. *knowledge relativity* aware version of Shannon's *information*

knowledge relativity does not require reference to any *object*, but it requires reference to *subjects*, as can be demonstrated on the example of a *knowledge relativity* aware version of Shannon's *information* (Fig. 6): A signal (the *representation* is assumed to already contain any channel distortion, S' in Fig. 3) sent by one *subject* could be *received* by a multitude of *subjects*. For each *receiver*, this signal could *represent* different *knowledge*, always in relation to each *receiver's knowledge* about the *sender* and each *receiver's intentions* for this signal. Thus each *receiver* could detect a different amount of *information* for the same signal, and the same signal would contribute to the *knowledge* of different *receivers* in different ways. For an individual *receiver*, Shannon's *information* definition as the degree of un-ability for predicting the next symbol (in relation to this *receiver's knowledge*) still holds. From a *knowledge relativity* aware perspective, *knowledge* is not contained in any of the involved *subjects*, nor can it be assembled from or reduced to *subject*-external components like those constituting symbolic *representations*: Since any understanding of the link between some specific *subject*, some specific *representation* and some specific *knowledge* always requires some other *subject* in the role of *evaluator*, the link between *representation* and *knowledge* will always be dynamic.

7 Conclusion

The main motivation for the introduction of *knowledge relativity* in this article was to facilitate the dialogue across disciplines, particularly between Informatics and Philosophy. But assuming a fertile dialogue and assuming this new notion proves useful, I would expect a direct effect both in Informatics and in Philosophy, and potentially in more disciplines. A fair amount of literature outside Informatics and Philosophy deals with issues similar to the ones exemplified above, and if it is true that we already live in a "knowledge society", some additional clarity on the nature of knowledge or at least the nature of the term *knowledge* should prove useful.

References

Gruber, T. R.: 1991, The Role of Common Ontology in Achieving Sharable, Reusable Knowledge Bases, *Principles of Knowledge Representation and Reasoning*, Morgan Kaufmann, Cambridge, USA.

Nake, F. and Grabowski, S.: 2001, Human-computer interaction viewed as pseudo-communication, *Knowledge Based Systems* **14**.

Newell, A.: 1982, The Knowledge Level, *Artificial Intelligence* **18**, 87–127.

Newell, A.: 1993, Reflections on the knowledge level, *Artificial Intelligence* **59**, 31–38.

Popper, K.: 1972, *Objective Knowledge: An Evolutionary Approach*, Oxford University Press, UK.

Shannon, C. E.: 1948, A mathematical theory of communication, *Bell System Technical Journal* **27**.

Stamper, R. K.: 1993, Social norms in requirements analysis, *in* M. Jirotka, J. Goguen and M. Bickerton (eds), *Requirements Engineering, Technical and Social Aspects*, Academic Press, New York, USA.

The Relationship between Metapattern in Knowledge Management as a Conceptual Model and Contragrammar as Conceptual Meaning

John D. Haynes (www.bus.ucf.edu/jhaynes/)

Department of Management Information Systems
University of central Florida
Orlando, Florida, USA

and

Pieter Wisse (www.wisse.cc)

Information Dynamics (Voorburg)
and University of Amsterdam
Netherlands

Abstract

In this paper we wish to examine a deeper issue underlying conceptual modeling, namely what could constitute conceptual meaning. Accordingly we have chosen to compare an analysis of "Metapattern" (Wisse, 2001), with "Contragrammar" (Haynes, 1999). We note that there are distinct similarities between Metapattern and Contragrammar at the stage of (1) an exchange of redundancy, and (2) flexibility through relative concepts. Conceptual modeling is an important aspect of Knowledge Management, and, clearly, as concepts vary with 'situations,' conceptual modeling must accommodate those variations. From this paper, it is hoped that, by examining features of Metapattern in relation to Contragrammar, we can thereby see why 'situations' entail meaning, and why Metapattern works as a model of behaviour, and why Contragrams can capture an essential feature of that 'situational' meaning.

Introduction

Metapattern (Wisse, 2001) is designed to encompass variety in agreement with various 'situations.' It essentially shifts atomicity to particular behaviors (Wisse, 2002b). Metapattern is sketched here with an emphasis on assumptions. Contragrammar (Haynes, 1999) is explicated via a shift from the human perception situations to a human conception of the meaning of that 'situation.' Firstly we shall examine Metapattern.

Metapattern: An exchange of redundancy

Social psychology instructs about the situational nature of behavior (Dewey, 1938; Carr, 1955). An actor, or agent, is assumed to reside in various situations. Hence variety exists in the actor's behavior. In fact, a particular behavior completely corresponds with the actor *as far as a particular situation goes*. Adding situation and behavior therefore turns inside-out the treatment of an actor as entity/object. Only an actor y's barest identity remains necessary and sufficient for relating a situation x and a behavior z (Figure 1.a). The whole of actor y is now reflected by his particular behaviors *across relevant situations* (Figure 1.b).

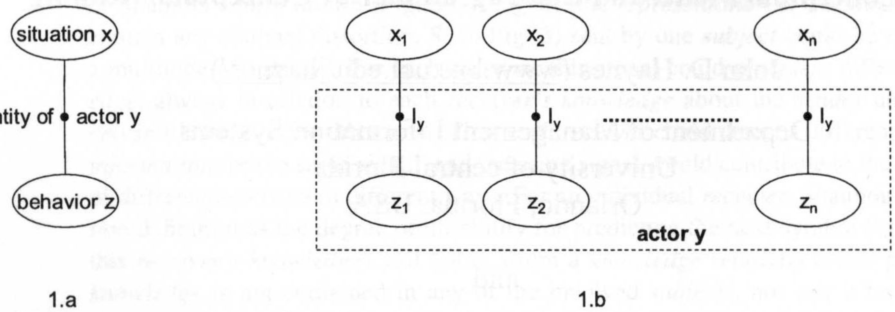


Figure 1: The conceptual triad of situation, (actor's) identity and behavior.

Please note that juxtaposition of behaviors rests on repeating the — reference to the — actor's identity. An instance of identity exists for every relevant situation, i.e. where the actor has a particular behavior. Through repeated identities, differences (particular behaviors) are reconciled with unity (one actor). In traditional modeling (entity-attribute-relationship, subject-predicate, object orientation, object-role, etc.), repetition of identity shows up as redundancy and is therefore ruled out. Metapattern, however, accepts repeating identities for situations as the deliberate price for avoiding hierarchical decomposition strictly within the boundaries of one actor itself.

Metapattern: Flexibility through relative concepts

Metapattern's compactness and flexibility comes from the assumption that situation, (actor's) identity and behavior are relative concepts. A rigorous set of modeling constructs applies throughout. Following the spatial orientation of Figure 1, decomposition can proceed both up and downward. Upward, for example situation x_1 can itself be considered as constituted by several actors' identities — presumably all different from I_y — appearing in a correspondingly less determined situation. Introducing levels, the original situation x_1 may be designated as situation $x_{m,\,1}$. The result of one step of upward decomposition is shown in Figure 2.

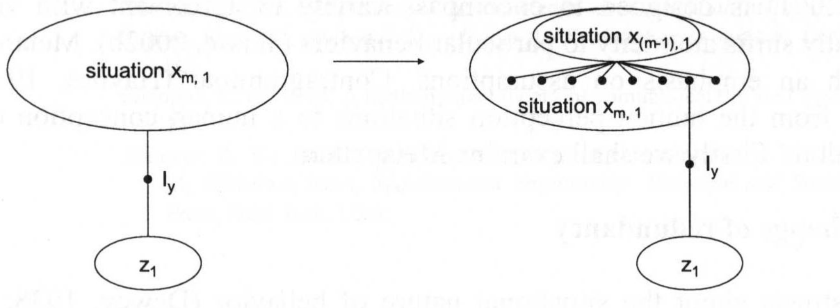


Figure 2: Upward decomposition.

In Figure 3, the right hand side of Figure 2 has been adapted to indicate what after one step of upward decomposition now count as situation, identity and behavior.

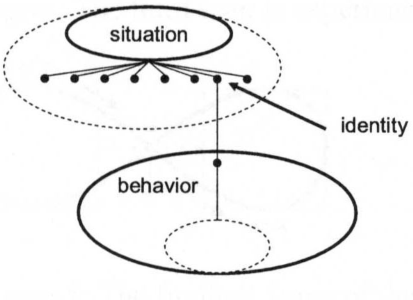


Figure 3: A shifted configuration of situation, identity and behavior.

Reading these figures in reverse order gives an impression of downward decomposition; the original identity is encapsulated within a more determined situation while the original behavior is decomposed into identities, each with situational behavior at a lower level of the conceptual model. There are no limits to upward and downward decomposition. A upper boundary condition must be formally set, however. It reflects the model's horizon, corresponding to the most generally accounted for, least determined situation. In metapattern's visual language (Wisse, 2001) it is a thick horizontal line. A conceptual model takes on the shape of a lattice of nodes. Some nodes are connected to the 'horizon.' Different instances of equal identity may be connected laterally to indicate that an actor's behavior in one situation is invoked from — his behavior in – another situation.

Accounting for (individual) differences in Knowledge Management

Metapattern readily supports management-of-knowledge where knowledge is considered the result of objective analysis. But, to what extent should knowledge management allow for human differences in daily practice? It is especially when account must be taken of essential differences that metapattern supports necessary and sufficient elaboration while maintaining cohesion. Does the experience of one employee provide the appropriate model of behavior for another employee? Is a report of an experience rightly called knowledge in the sense of useful? Should conditions of usefulness be made explicit?
It might be that the careless assumption of equality, similarity, and so on, between situations is risky. In such cases, it is recommended that situations are explicitly classified with regard for individual involvement. This requires the — identity of the — reporter to be stated, too, as part of the situation of the experience-turned-report for knowledge management. In addition, the reporter's motives might be relevant, suggesting a perspectival turn (see below).
Metapattern suggests an actor invokes a particular behavior as befitting a particular situation. The patterns an actor is primarily recognizing are therefore situations. Recognition is not merely passive, though. An actor, artificial or not, would be overwhelmed when it lacks a pattern of its own to establish situations-as-patterns. Traditionally, an actor's self-directed activity is attributed to goals. A goal is seen as different in kind from, for example, a behavioral specification that should be executed to accomplish it.

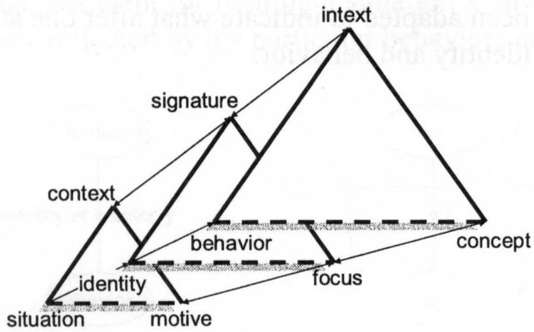


Figure 4: Semiotic ennead.

An extension of the semiotic triad (Peirce, 1906) into an ennead (Wisse, 2002a) argues for a relative status of goal, or motive. In Figure 4, the term 'motive' is preferred. In semiosis, an actor constructs a sign from a configuration of — again, the relative concepts of — motive, focus and concept. In short, he brings a perspective to his own observational behavior. The resulting sign is transformed in cycles of semiosis into another configuration of motive, focus and concept. That is, semiosis changes the actor's perspective, resulting in one particular behavior or another.

Contragrammar: An exchange of redundancy

In this section a consistent notation is presented for the movement from perception to conception: from the perception of content (implicit stage) through to the experience (or re-experience) of forming a conception (explicit stage).[1] The exchange of redundancy occurs at the level of perception, until the mode of conception is reached.

Consider the following statement: The culture of consumption is the consumption of culture[2]. The point being made is that a culture ideally emerges out of non-selfish motives. It arises out of the common good and common interests of a group of people. If the group of people turn on themselves and begin to treat each other as objects to be "consumed" rather than as individuals deserving respect, then the original culture will begin to disintegrate. Similarly if a culture naturally emerged out of common interests that were nourished and supported by a group of individuals began to be exploited and "consumed" by the individuals, rather than being supported in an environment of learning, then again the culture would begin to disintegrate.

The above statement has the following structure: The X of Y is the Y of X, labeled by Haynes (Haynes, 1999) as a Contragram. The structure performs its function as an idea in a holistic way. In other words, the parts, the X and Y only function as the contents of a wave-form. They are the X and Y drops in the oceanic wave-form of the Contragram. It is interesting to note that Peterson makes the comment that interpreting wave-forms in mathematics in relation to "infinitely many measurements with infinite precision" in the context of how "two different objects appear alike in every measurable way" is still yet to deal with tough problems. (Peterson, p 101, 1998).

[1] This notation was first presented by Haynes in 1990 and published in 1991. See (Haynes, 1991). See also (Haynes, 1999).

[2] Contributed by B.C.Birchall.

In the comprehension of this particular Contragram we begin with the perception of (the) X of Y and continue to the Y of X. It is only in the return to the X of Y that the perception-in-gestalt begins. The third scan is experienced as the whole Contragram. See figure 5.

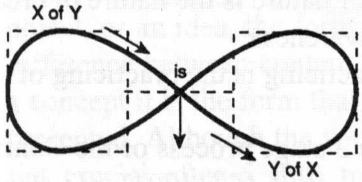


Figure 5: The Implicit Stage of the Contragram: perception (of content).

The notion of the wave-form is all the more reinforced by the fact that X and Y cannot be substituted by words that represent perceivable things. For example, consumption is meant in a metaphorical sense; culture is elusive, but is nevertheless real.
The process can be presented as follows:

(a) At the level of perception: the terms (in lowercase) are represented as the x of y is the y of x. At this level no conception is apparent. The terms, indicated in lowercase, are functioning as words and no more than that.
(b) The reader now completes the scan of perception. It may be that one term is recognised for its conceptual capacity in which case the Contragram becomes, notationally, the X of y is the y of X.
(c) On the third scan a shift to conception is made and the Contragram is recognised as meaningful, and the notation becomes the X of Y is the Y of X.

Further analysis of the above process reveals that on the surface of the Contragram (at the level of perception), there is redundancy in "the x of y is the y of x." Secondly, there is certainly, on the surface, ambiguity, that is, how is x of y apparently differentiated from the X of Y? But if the shift is made to conceiving then the ambiguity is (like the launch pad) left behind. It is left behind because redundancy is found in objects to be logically dissected or analysed, and not in concepts. This letting go then allows the apparent ambiguity to diffuse itself in a shift in seeing that takes in a gestalt (the states of which are neither 'parts' that repeat, but are simply contained in that flux of the wave-form). This can be regarded as a discourse that shifts from death, that is, the death of the object-as-perceived to the life-of-the-gestalt as conceived. A conception which returns again to reconstitute or resurrect that death as Life, which is to say, to resurrect that death as meaning, or in other words, gives meaning to that apparently ambiguous content. This is the sense by which Contragrammar is a consistent notation for the enunciation or resurrection of meaning. The fact that Contragrammar is non-logical does not entail that it is illogical.

Contragrammar: Flexibility through relative concepts

The Contragram makes sense by succeeding to bring together apparently conflicting relative terms in the form of a shift to conception. Once the shift is complete apparent conflict is resolved. And it is the gestalt of all terms together, which in their being brought together (the journey of the return-to-the-beginning) either metamorphoses their content into form (thus giving the Contragram meaning; enabling its be-ing a Contragram), or the terms are 'frozen' at the level of content, in which case we have no Contragram. Or the 'shift' is possible, but the conception (on the part of the reader) is lacking. Thus the Contragram is a pointer to a sense of form, which has the following apparently contradictory or paradoxical presentation:

Further Examples of Contragrams[3]:

Here are two Contragrams by John Haynes: (1) The creation of nature is the nature of creation. (2) The awareness of consciousness is the consciousness of awareness.

And two Contragrams by Pieter Wisse: (1) The mastery of practicing is the practicing of mastery. (2) The nature of nurturing is the nurturing of nature.

It may be of assistance to briefly explain Contragrammar (the 3 stage process of the Contragram philosophically as follows (paraphased from Haynes, (Haynes, page 157, 1991), see also figure 6). In the two-part Contragram, the terms 'X' and "Y' must be co-extensive. The Contragram is a 'logic' of universals; a 'logic' of meaning; a 'logic' of the subject becoming meaning, i.e. a 'logic' in which the subject/predicate distinction is overcome, and in that overcoming, becomes a Real term. Whereas, the logic of either/or (traditional logic, symbolic logic), will always particularise the predicate through the subject. For example, assume we are looking at a certain table which is square. So we can assert: the table is square. And we could conceive of its contradictory: the table is not square. But squareness is not tableness even though squareness happens to be true of this particular table. In the Contragram, we have no subject, we have instead a 'non-logic' of the concept, implicitly (at the level of perception) in the x of y is y of x, becoming explicit in X of Y is the Y of X. In this case, the X of Y is the Y of X is not a fact (i.e. it does not exist independently of its being conceived) but, instead, is its being conceived (i.e. what is known as its being known). Some person, Z, perceives that S is P, and here we have two distinct facts:

(a) Z is perceiving that S is P, which is either true or false;

(b) S is P, which is either true or false.

But when Z conceives S As P, we have one fact: (And herein lies the key to comprehending and creating new Contragrams: the "X of Y" is conceived of as "X As Y", not the as of comparison, but the As of revelation (see Haynes 1999c)).

Z is conceiving S AS P, which is either true or false.

Which Contragrammatically, is expressed as follows:

Z conceives (X of Y AS Y of X), and this is either true, i.e.

[Z conceives (X of Y AS Y of X)] or false, i.e. NOT [Z conceives (X of Y AS Y of X)].

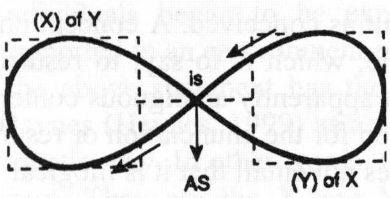


Figure 6: The Explicit Stage of the Contragram – arriving at meaning: Conception (or form).

[3] For a more thorough treatment (and examples) of Contragrams see Haynes (Haynes 1999).

Convergence of Metapattern and Contragrammar: the Concept

From Contragrammar we have noted that when we perceive an object the object necessarily needs to be in view, or at the very least presented for the perception process, and in this respect the content of the object is a requirement. On the other hand, when we conceive of an object, or an idea, the form or forming of an object or idea is of paramount importance. This difference between content and form is central to the notion of a concept. In comprehending a concept it is the form that we are apprehending, rather than the way in which the concept is presented. Although the way in which the concept is presented is also crucial. However, it is not crucial after a shift to the concept (idea) itself has been made. Let us say that the presentation of the concept or central idea is like a launch pad, which is critical for launching, but not directly related to the comprehension of the space-craft once the space-craft has been launched (once the idea has been grasped). For example, in mathematics if we are presented with a solution to some problem we firstly perceive the equations or mathematical proof but to conceive of the central idea or theory that the equations are capturing we have to experience or re-experience the idea that the equations are pointing to. Similarly in philosophy, the words, which capture a certain point of view, are initially merely perceived. The initial perception then gives way to, or changes into, a state of conception as the reader comes to a comprehension of the meaning (or at the very least validity) of what is being presented or read. From our analysis of Metapattern, we can see why formal knowledge management is worth substantial efforts. Suppose, when documenting an experience, a reporter is asked to include more information than he can realistically be expected to provide. That would immediately make using what was nonetheless recorded as knowledge extremely risky. Wouldn't a superior alternative not be a simple personal reference, to be called upon only when a particular need arose? Conceptual models should assist making practical judgements before incurring the bulk of costs. Again, Metapattern doesn't impose any conceptual limits upon modeling knowledge even when individual differences must be included. The relative nature of its key concepts, situation, identity and behavior, guarantee open-ended opportunities for decomposition both upward and downward. Precisely because conditions for rigor are secure, the question of relevance deserves priority. This is precisely why, for the present, Metapattern is a robust example of the essential inner conceptual workings of Contragrammar, yet is able to apply in real life situations without imposing any conceptual limitations which clearly Contragrammar does in terms of the peculiar nature of the X and Y terms used. In the following consideration, we refer to bi-contragrams (2 terms) as explicated above, and poly-contragrams entailing multiple terms using Metapattern itself to illustrate how Contragrammar could be extended from 2 terms to multiple terms. This will be presented during the workshop session for this paper.

Conclusion

Given the redundancy similarities between Metapattern and Contragrammar and how each can achieve flexibility through relative concepts we gain a sense of how Contragrammar is a conceptual way of explaining some important aspects of Metapattern. As knowledge management advances to cover ever more complex human endeavors it inevitably needs to acknowledge what constitutes conceptual meaning. As we have seen from an explication of Metapattern, the nature of the challenge is similar to empowering an artificial actor with behavioral variety. But inevitably if Knowledge Management is to effectively – meaningfully - move to Conceptual Management, a thorough understanding of what constitutes the movement from perception to conception as identified by the Contragram must be embraced more fully in Conceptual Modeling.

References

Carr, L.J. *Analytical Sociology: Social Situations and Social Problems*, Harper, New York (USA), 1955.
Dewey, J., *Logic, The Theory of Inquiry*, Holt, Rinehart and Winston, New York (USA), 1960, originally published in 1938.
Haynes, J.D. The Phenomenology of Neural Networks, in: *Proceedings of the 1990 Conference of the Australasian Association for Phenomenology and Social Philosophy: Hermeneutics, Phenomenology & Technology*, pp149-160, edited by J. Dunphy, P.Bilimoria, Deakin University Press (Victoria, Australia), 1991.
——— *Meaning As Perspective: The Contragram*, Thisone (New Zealand), 1999.
——— *Perspectival Thinking: for Inquiring Organizations*, ThisOne and Company, Palmerston North (New Zealand), 2000.
Peirce, C.S. Pragmatism in retrospect: a last formulation, in: *Philosophical writings of Peirce*, edited by J. Buchler, Dover, New York (USA), 1955, essay originally written in 1906.
Peterson, Ivars, *The Jungles of Randomness – Mathematics at the Edge of Certainty*, Penguin Books (UK), 1998.
Wisse, P.E. *Metapattern: context and time in information models*, Addison-Wesley, Boston (USA), 2001.
——— *Semiosis & Sign Exchange: design for a subjective situationism, including conceptual grounds for business information modeling*, Information Dynamics, Voorburg (Netherlands), 2002a.
——— The ontological atom of behavior, in: *PrimaVera*, working paper nr 2002-06, Amsterdam University, Amsterdam (Netherlands), http://primavera.fee.uva.nl/, see publications/working papers, 2002b.

About the authors:

John D. Haynes (www.bus.ucf.edu/jhaynes) is a Long-Term Visiting Full Professor in the Management Information Systems Department at the University of Central Florida, Orlando, Florida, USA. John holds a BCom (Computer Science) and BA (Philosophy) from Newcastle University in Australia, an MA (Philosophy) from Macquarie University in Australia, and a PhD (Information Systems, Artificial Intelligence, Philosophy) from Bond University in Australia.

Pieter Wisse (www.wisse.cc) is the founder and president of Information Dynamics, an independent company operating from the Netherlands and involved in research & development of complex information systems. He holds an engineering degree (mathematics and information management) from Delft University of Technology and a PhD (information management) from the University of Amsterdam. At the latter university, Pieter is affiliated with the PrimaVera research program in information management.

Transparente Schichten - Perspektiven der Informatisierung des Wissens

Gerhard Gamm, Stephan Körnig

Technische Universität Darmstadt, Institut für Philosophie, Residenzschloss, 64287 Darmstadt
gamm@phil.tu-darmstadt.de, koernig@ifs.tu-darmstadt.de

Abstract. Ausgehend von der aus unserer Sicht höchst problematischen Unterscheidung zwischen propositionalem und nicht-propositionalem Denken, befassen wir uns mit ähnlich gelagerten Unterscheidungen, die bei der Modellierung von Wissen durch Informationssysteme in Anspruch genommen werden. Hierbei wird insbesondere die Unterscheidung zwischen explizitem und implizitem Wissen kritisiert, sofern sie von unzutreffenden Annahmen über den Erwerb, den Umgang, die Weitergabe oder die sinnvolle Nutzung von Wissen ausgeht. Die sog. „Semantic Web" Initiative, die auf Tim Berners-Lee (dem „Erfinder" des Internet) zurückgeht und gegenwärtig einen Teil konkreter Standardisierungsaktivitäten des W3C (Konsortium des World Wide Web) ausmacht, wird in diesen Zusammenhang gestellt und diskutiert. Das Interesse an den sog. „Ontologien", die als ein vielversprechendes Werkzeug für die Integration von Informationssystemen auf semantischer Ebene angesehen werden, ist aus unserer Sicht begrüßenswert - es wird aber darauf hingewiesen, dass solche Instrumente auf Kontexte angewiesen sind, die prinzipiell nicht durch sie selbst bereitgestellt werden können. Anhand des Begriffs der Transparenz wird diese Kontextualisierung dargestellt und es werden Forderungen formuliert, die aus unserer Sicht für eine gelingende Praxis des Wissensmanagements unabdingbar sind.

1. Einführung - Implizites Wissen und seine Kontexte

In einem bemerkenswerten Aufsatz *Ein Rahmen für die Wissenspräentation (A frame work for representing knowledge)* [1] geht Minsky der Frage nach, wie Rahmen, auf die wir uns in unserem Alltagsleben und -handeln stützen, von Computerprogrammen bearbeitet werden, wie Rahmen, Kontexte, Hintergrundwissen, Frames of Reference, Perspektiven usf. auf solche für informationsverarbeitenden Systeme kompatible Wissensstrategien „heruntergebrochen" werden können: Wie können Kontexte von Computern repräsentiert werden? Wie sind Rahmen oder Kontexte in Softwarearchitekturen integrierbar? Wie lassen sie sich in Computerprogramme übersetzen? Minskys Beispiel ist unter anderem der Rahmen oder das Hintergrundwissen, das ein Kind aktivieren muss, wenn ein anderes Kind Geburtstag hat und es ihm ein Geschenk machen will: Welches Rahmenwissen muss man aktualisieren, um Motive, Handlungen Reaktionen richtig einzuschätzen, wenn zum Beispiel das Kind, das ein

Geschenk gekauft hat, von einem dritten, etwa einem anderen Kind hört, das betreffende Kind hätte das Geschenk schon, einen Drachen beispielsweise. Es geht darum, die impliziten Voraussetzungen des Wissens, die wir wie selbstverständlich machen, zu benennen, zum Beispiel die sozialen Präsuppositionen zu explizieren, die man für ein angemessenes Verständnis einer Geschichte bzw. eines Sprachspiels oder der folgenden Sätze braucht.

„Er hat bereits einen Drachen. Er möchte, dass Du ihn zurücknimmst." Minsky fährt fort: „Doch welchen Drachen soll sie zurücknehmen? Wir wollen sicher nicht, dass Petra Peters alten Drachen zurücknimmt. Um den Bezugsgegenstand des Pronomens ‚ihn' zu bestimmen, muss man eine Menge über das vorausgesetzte Szenario wissen. Sicher bezieht sich ‚ihn' auf den vorgeschlagenen *neuen* Drachen. Woher weiss man das aber? (Man beachte, dass jede einzelne Erklärung für sich unzureichend sein kann). Im Allgemeinen beziehen sich Pronomen auf die zuletzt erwähnten Gegenstände, dieses Beispiel zeigt jedoch, dass der Bezugsgegenstand von mehr als der lokalen Syntax abhängt. Nehmen für einen Augenblick mal an, dass wir versuchen, den Standardunterrahmen ‚ein Geschenk kaufen' anzuwenden." (Minsky 101, 102).

Uns geht es im Augenblick nicht darum, Minskys Voraussetzungsanalyse des Wissens zu diskutieren - es geht zunächst um die Frage, nach der Gesamtstrategie, die Minsky und ein Teil der KI-Forschung verfolgen, wenn sie Kontext und Rahmenwissen als implizites Wissen konzeptualisieren, das man, in welcher Form auch immer, explizit machen muss. Das heisst, man verwandelt das implizite Wissen in ein gegenständliches Wissen, oder besser: man transformiert, überträgt es in ein Wissen, das eine propositionale Form hat. Man sucht sich des Rahmenwissens zu vergewissern, indem man es in allgemeine Standardannahmen (Standardrahmen) übersetzt. Diese Standardannahmen sollen gleichsam Standardsituationen widerspiegeln, Situationen also, die als typisch bzw. prototypisch gelten können. Kurz, die Frage lautet, lässt sich Kontextwissen nach Art von Sachverhaltswissen errechnen, das heißt als vorher festgelegte Auswahl von Objekten, Eigenschaften und Relationen? Sind Kontexte gewöhnliche „Gegenstände"? Gehorchen sie üblichen gegenstandsanalogen Regeln? Sind Kontexte Gegenstände herkömmlicher Art in der Weise, dass sie propositionalisierbar sind, dass sie in Standardannahmen übersetzt werden können?

Unsere Fragestellung zielt also keineswegs darauf, zu klären, ob Computerprogramme prinzipiell in der Lage sind, Wissen zu prozessieren oder zu modellieren. Es geht vielmehr darum, zu fragen, ob die Formalisierung von Wissen nicht zwingend in einen unendlichen Regress gerät, der dadurch entsteht, dass die Unterspezifiziertheit von Bestimmungen durch eine genauere Bestimmung des Kontextes eingeholt werden soll. So hat etwa Wittgenstein in seinen „Philosophischen Untersuchungen" dargelegt, dass die „unanalysierte Form" von Sätzen im Rahmen des Alltagswissens und -handelns eine Anschlussfähigkeit besitzen kann, die durch eine weitere Analyse entweder verloren geht oder nicht weiter „erhellt" werden kann[2].

2. Standardisierung und Werkzeuge zur Modellierung von Wissen

An dieser Stelle ist es vielleicht ganz nützlich, einen Blick auf konkrete Anstrengungen in einem - so scheint es zumindest - weniger ambitionierten Bereich zu werfen, als den, in dem die KI-Forschung ihre Erfolge erzielen will. Es geht hierbei um die „Semantic Web“ Initiative zur Integration von Diensten und Angeboten im Internet. Schon hier wird sich zeigen, dass bei der Informatisierung des Wissens Kontexte als nicht-formalisierbare Bedingungen des Informationsaustausch präsent sind und dass die Strategie, diese in einem propositionalen Netzwerk „aller implizierten Propositionen“ einfangen zu wollen, aus prinzipiellen Gründen zum Scheitern verurteilt ist. Dies verstehen wir allerdings nicht so, dass die Informatisierung des Wissens nicht möglich wäre. Aus unserer Sicht wäre eine solche philosophische Position ohnehin belanglos: da diese Informatisierung schon längst große Teile des gesellschaftlichen Wissens durchdringt und fortbestimmt, gilt es vielmehr zu verstehen, unter welchen Bedingungen nun etwas „gewusst“ werden kann und in welcher Weise die hierbei in Anspruch genommenen Wissenstechnologien selbst einen Kontext für das Wissen geben.

Vision des „Semantic Web“

Bei der Semantic Web Initiative geht es z.Zt. vor allem darum, die Integration von Informationssystemen zu automatisieren. An Geschäftsvorgängen im Internet können unterschiedliche Partner beteiligt sein (Banken, Abrechnungssysteme, Lieferanten, Suchdienste mit Verfügbarkeitsrecherche usf.). Im Hintergrund z.B. einer einfachen Internetbestellung müssen daher mitunter Dutzende unterschiedliche Informationssysteme miteinander kommunizieren, die intern völlig unterschiedlich organisiert sein können. Um so „offener“ ein solches System ist, d.h. um so leichter es für verschiedene Anbieter ist, in diesem „Konzert“ mitzuspielen, um so attraktiver werden die Angebote für den Kunden sein. Zugleich fällt damit eine Abstimmung der Informationssysteme „von Fall zu Fall“ aus. Es wird, so paradox es klingt, ein Standard benötigt, der die Offenheit des Informationsaustausches gewährleistet.

Das in diesem Bereich bedeutendste Standardisierungsgremium ist das sog. W3C-Konsortium, da es für die im World Wide Web benötigten „Beschreibungssprachen“ zuständig ist. Diese Sprachen reichen vom bekannten HTML, mit dem der Inhalt und die Gestaltung von Web-Seiten festgelegt werden können, bis hin zu WebOnt, das als ein allgemeiner „Rahmen“ für die Beschreibung von Konzepten, den sog. „Ontologien“[1], zur Verfügung gestellt wird.[2]

Tim Berners Lee, der Direktor des W3C-Konsortiums, war zu Beginn der Arbeit dieses Konsortiums von einer weitaus weitreichenderen Zielsetzung - die man mit Recht eine Vision nennen kann - überzeugt. Er vertrat die Auffassung, das Web könne sich selbst zu einem gigantischen semantischen Netzwerk entwickeln, also zu einem globalen Wissensspeicher - verbunden mit der demokratischen Zielsetzung eines „glei-

---

[1] Diese Begriffswahl mag manchen problematisch erscheinen. Aus Sicht der Philosophie impliziert ein solcher Begriff jedenfalls weitaus mehr als das, was diese sog. Ontologien leisten. Diese dienen der Schematisierung von Konzepten, die bei der Strukturierung von Informationen zugrunde gelegt werden. Ontologien im philosophischen Verstande zielen dagegen auf eine (kategorial verfasste) Ordnung des Seins. Sie können deshalb auch Gegenstand der Kritik sein - etwa indem das kategoriale Verfasstsein von Welt/Wirklichkeit bestritten wird. Die Konzeptualisierung von Informationsräumen kann dagegen nur mehr oder weniger „zweckdienlich“ sein.

[2] Dieses Spektrum kann natürlich auch ganz anders „aufgespannt“ dargestellt werden. Unter www.w3.org kann man sich einen Eindruck von der Fülle der Standards verschaffen.

chen Zugangs“ für alle. Wie wir mittlerweile wissen: Sowohl die Nutzbarkeit des Web als auch die reale Nutzung sind von solchen Vorstellungen weit entfernt. Neben der kommerziellen Besetzung des Internet - die durchaus auch zur Attraktivität des Web beiträgt - hat sich das Web als ein Schauplatz für Nebensächliches, Selbststilisierungen, Lustgewinnen usf. etabliert. Informationssuchende werden mit einer Heterogenität konfrontiert, die sich nur mittels selektiver Filter (z.B. Google, Bookmarks, Portale) auf ein erträgliches Maß reduzieren läßt. Wenn wir das Web dennoch als einen Informationsraum ansehen und bei der Suche nach Informationen alle möglichen verfügbaren und relevanten Angebote berücksichtigen wollen, dann stehen wir bekanntlich vor dem unlösbaren Problem einer vollständigen Webrecherche.

Die aktuelle „Semantic Web“ Aktionslinie des W3C erscheint, vor diesem Hintergrund gesehen, mehr als das Eingeständnis gescheiterter Ambitionen, denn als die Fortschreibung der ursprünglichen Vision. Es wird nun eine Infrastrukturinitiative beschrieben, die spätere Zielsetzungen ermöglichen soll. Für die auf diese Infrastruktur später aufsetzenden Nutzungsformen weithin *unsichtbar* soll eine Schicht der Maschine-zu-Maschine-Kommunikation etabliert werden.

„The Semantic Web is a vision: the idea of having data on the Web defined and linked in a way that it can be used by machines not just for display purposes, but for automation, integration and reuse of data across variant applications.

In order to make this vision a reality for the Web, supporting standards, technologies and policies must be designed to enable machines to make more sense of the Web, with the result of making the Web more useful for humans.“ (W3C Semantic Web activity statement)

Das maschinenfreundliche Web steht nun im Zentrum der Initiative, auf dass das Web letztendlich für den Menschen nützlich werde. Man kann hier durchaus von einer *Verkehrung der Vision* sprechen - insbesondere dann, wenn der Verdacht besteht, dass schon die Schaffung dieser Infrastruktur auf hartnäckige Probleme stößt, die noch für Jahrzehnte Arbeitsfelder bestimmen werden. Diese Probleme sind keineswegs neu oder unerwartet, haben sie doch mit dem Unterschied von Syntax und Semantik zu tun, der bei jeder Phase des Aufbaus von Informationssystemen eine Rolle spielt.

Die Modellierungsaufgabe bei der Anwendungsentwicklung

Die „implizite Semantik“ von Informationssystemen wird in der Systemtheorie mit der Unterscheidung von Codierung und Programmierung thematisiert. Für die Codierung bestimmter Sachverhalte benötigt die Programmierung Kriterien. Solche Kriterien können z.B. an einem Datenmodell einer Anwendung abgelesen werden. Das Datenmodell zeigt dem Experten auf, wie die fachlichen Probleme (etwa die Verwaltung von Versicherungspolicen) in Datenstrukturen übersetzt werden und welche Abhängigkeiten zwischen den Daten durch dieses Modell berücksichtigt werden. Abgesehen davon, dass zwischen Modellbildung, der fachlichen Beschreibung und der Praxis zahlreiche Unterschiede bestehen, die meist erst bei der „Abnahme“ des Programmsystems deutlich hervortreten, ist schon die korrekte Überführung der durch die Ist-Analyse ermittelte Semantik in entsprechende Datenstrukturen nicht immer zu garantieren. Weitere Kriterien, die in einem statischen Datenmodell nicht abgebildet werden können, werden durch den Code selbst erfüllt, der für eine korrekte Abbildung von

Sachverhalten aus dem Datenmodell heraus zu sorgen hat - bzw. für eine korrekte Encodierung von Sachverhalten im Datenmodell.

In der Entwicklung der Programmiersprachen und der dazugehörigen Modellierungstechniken hat sich diese Problematik ebenso niedergeschlagen.

- Schon die so genannte „strukturierte Programmierung“ befaßte sich mit dem Unterschied zwischen Codierung und Programmierung. Sie stellt vor allem die Wartbarkeit von Programmcodes sicher, indem sie die Programmierlogik, nach der ein Programmierer vorgegangen ist, selbst explizit macht. Programmierer benutzen dann Struktogramme, die unabhängig von der verwendeten Programmiersprache als eine Art lingua franca die Zusammenarbeit von Teams erleichtern. Ebenso wird das Konzept der Wiederverwendbarkeit und der Modularität von Programmen unterstützt. Die Aufgabe der fachlichen Modellierung - die Semantik der fachlichen Probleme in die syntaktischen Konstrukte und die konkreten Entitäten des Datenmodells zu überführen, wird hier allerdings noch nicht angegangen.
- Im Bereich der Datenmodellierung kann die Entity-Relationship Diagrammtechnik als eine Ergänzung angesehen werden.[3] Die Datenzusammenhänge werden (zumindest in den neueren Varianten dieser Technik) im Kontext der fachlichen Modellierung selbst dargestellt - und sie kann sich hierbei auf das relationale Datenmodell stützen, das zum konkreten Aufbau einer Datenbank führt.
- Im Bereich der objektorientierten Programmierung schließlich werden durch eine umfangreiche Diagrammtechnik unterschiedliche Aspekte der Semantik berücksichtigt. Taxonomien, Anwendungsfälle und Interaktionen werden durch eine standardisierte Diagrammtechnik beschrieben und können als Blaupausen bei der Anwendungsentwicklung genutzt werden. Zugleich ist mit dieser Programmiertechnik vollends deutlich geworden, dass es eine vollständige formalisierte Darstellungstechnik der „impliziten Semantik“ nicht gibt. Unterschiedliche Aspekte der Modellierung (so z.B. das Typsystem im Unterschied zum Komponentenmodell) werden als jeweils unterschiedliche Sichten festgehalten - die Designer selbst müssen diese Sichten in ihren Köpfen zusammenbringen, vor allem dann, wenn es um Weiterentwicklung des Designs, Erkennen von Schwächen im Konzept usf. geht. Darüber hinaus wird im Bereich des objektorientierten Designs gefordert, dass bei der Entwicklung kontinuierlich Personen eingebunden sind, die das fachliche Knowhow aus Sicht der Anwender in den Gesamtprozeß einbringen. Auch dies aus der Erfah-

[3] Was hier als „Phasen“ einer Entwicklung dargestellt wird, ist im Kontext realer Anwendungsentwicklung nicht so einfach zu trennen. So sieht z.B. Heide Balzert die Entity-Relationship Diagrammtechnik als Teil der „strukturierten bzw. klassischen Entwicklung“ [3]. Hier sei ein Auseinanderfallen der unterschiedlichen Modellierungsmethoden zu konstatieren, während in der aktuellen objektorientierten Entwicklung dieselben Konzepte durchgängig benutzt werden könnten. Auch dem entspricht die aktuelle Entwicklung nicht ganz - vor kurzem wurde die neue Spezifikation des Standards der objektorientieren Modellbeschreibungssprache „Unified Modeling Language“ herausgegeben (UML 2.0) und diese lässt die Diagrammtechnik des Struktogramms wieder auferstehen. Die Entwicklung neuer Techniken ersetzt also nicht umstandslos die älteren - die neuen Techniken zielen vielmehr auf die Lösung neuer Probleme, die vor allem mit der zunehmenden Komplexität der Softwareentwicklung zu tun haben.

rung heraus, dass Design und Spezifikation der zu erbringenden Funktionen sich gegenseitig bedingen.

Angesichts dieser unterschiedlichen Anstrengungen, die noch keineswegs zu einem Abschluss gekommen sind, wird deutlich, dass die Aufgabe, fachliche Probleme durch ein in seinen formalen Eigenschaften beschreibbares System abzubilden, selbst nicht vollständig formalisierbar ist. Die Einbindung in eine soziale Praxis bildet letztendlich den Prüfstein für ein gelungenes Design. Derartige, sich beständig in Revision befindliche Informationssysteme über gemeinsame Schnittstellen miteinander zu vernetzten, ist eine Aufgabe, welche die genannten Schwierigkeiten vervielfacht (und nicht wenige Softwarefirmen in den Ruin getrieben hat). Der Ansatz der Semantic Web Initiative, so wie er gegenwärtig formuliert wird, ist zudem ein sog. *generischer* Ansatz. Es sollen nicht nur im Voraus bekannte Systeme miteinander verknüpft werden (so dass Transferkomponenten für den Informationsaustausch von Systemen mit unterschiedlicher internen Repräsentation (Datenmodell) entwickelt werden), vielmehr soll eine allgemeine Transferkomponente standardisiert - also unabhängig von den konkreten Eigenschaften einzelner Systeme spezifiziert - werden. Diese Komponente müsste, damit überhaupt eine Aussicht auf Erfüllung einer derartigen Transferleistung besteht, über das verfügen, was wir oben „Rahmenwissen" genannt haben. Sie müsste für den Austausch von Informationen zwischen Systemen, die Informationen für unterschiedliche Zwecke auch unterschiedlich repräsentieren bzw. modellieren, auf Strukturinformationen (etwa das Datenmodell) zugreifen können und diese im Hinblick auf die Aufgabenstellung des Transfers interpretieren.

An dieser Stelle geht also erneut darum, implizites Wissen zu explizieren. Das konkrete Instrument, mit dem dies geschieht, sind Sprachen, die Aufbauinformationen über Informationen enthalten (Schemainformationen) und die es ermöglichen, die Konzepte (auf die das Datenmodell bezogen ist) auszudrücken. Diese Sprachen werden, wie oben erwähnt, „Ontologien" genannt.

2.3. Ontologien und ihre Reichweiten

Die immer wieder zitierte Definition von Ontologien stammt von T.R. Gruber aus dem Jahre 1993 [4]: *Eine Ontologie ist eine formale und explixite Spezifikation einer geteilten Konzeptualisierung*. Unter Konzeptualisierung wird hierbei ein abstraktes Modell eines bestimmten Weltausschnitts verstanden, das die für diesen Ausschnitt relevanten Phänome identifiziert. Formal und explizit bedeutet in diesem Zusammenhang, dass es eine syntaktische Repräsentation dieser Konzeptualisierung in einer formalen Sprache gibt: definiert werden die Klassen oder Typen von Objekten, Relationen, Abhängigkeiten und Regeln des logischen Schließens (Ableitungsregeln).[4]

---

[4] Man spricht von *light weight ontologies*, wenn diese in der Hauptsache ein Schema bereitstellen, das zur Strukturierung eines Konzeptes eingesetzt wird. Ein solches Schema bietet eine Begriffstaxonomie, Attributs- und Relationsdefinitionen. Ein Thesaurus, ein Glossar oder ein Entity-Relationship-Diagramm können einer leichtgewichtigen Ontologie gegenüber durchaus als gleichwertig angesehen werden.Kommt zum Schema noch die Möglichkeit hinzu, prädikatenlogische Formeln auf Ausdrücke anzuwenden, dann

Wenn wir nun diese Definition auf den genannten Problembereich beziehen - den Transfer von Informationen, für den die jeweils unterschiedlichen semantischen Kontexte interpretiert werden müssen - dann sehen wir einen Widerspruch. Ontologien, die als Instrumente der Integration eingesetzt werden sollen, müssen domänenspezifisch sein.[5] Ihre Inanspruchnahme für eine universeile oder generische Transferkomponente könnte nur unter der Verkennung dieser prinzipiellen Einschränkung stattfinden. Die gewählten Ausdrucksmittel einer Ontologie ergeben sich - wie aus dem obigen Abschnitt über Modellierung deutlich geworden sein sollte - nicht aus dem Problembereich selbst. Die „Taxonomien" einer Ontologie können gut oder schlecht designt sein - ebenso wie die Relationen zwischen den Typen oder Klassen einer Ontologie. Diese syntaktischen Elemente der Ontologie sind mit hohem Aufwand formulierte explizite Annahmen über Zusammenhänge in einem Problembereich - und müssen sich in konkreten fachlichen Zusammenhängen bewähren. Kurz gesagt: Sofern Informationssysteme einer „gemeinsamen Domäne" zuzuordnen sind, können ihre Strukturen über eine solche Ontologie integriert werden. Diese Transferleistung entspricht damit aber nichts anderem als der Standardisierung von Strukturen im Aufbau von Informationssystemen, die vergleichbare fachliche Probleme lösen sollen. Ihr „Rahmenwissen" vermag sich nicht von den Abstraktionen abzulösen, die ihrem Design zu Grunde liegen.

3. Forderung nach Transparenz

Das Wort „Transparenz" wird im technischen und im wissenschaftlichen Kontext völlig unterschiedlich gebraucht. Ein Element eines Systems (z.B. ein Protokoll oder eine Verarbeitungsstufe) ist im technischen Sinne „transparent", wenn es vom Benutzer weder gesehen wird noch eine Kenntnis über es für die Nutzung benötigt wird. Dies stellt u.U. eine erhebliche Entlastung für den Nutzer dar. Bei der Aufbereitung von Information in wissenschaftlichen Zusammenhängen ist eine solche Transparenz nicht immer wünschenswert.

Der mit dem Begriff Wissen verbundene Erkenntnisanspruch dagegen enthält Transparenz als eine Forderung. Wer etwas weiß, muss sichtbar machen können, auf welcher Grundlage die Erkenntnisse erzielt wurden. Rechenschaft muss nicht nur für das abgelegt werden, was behauptet wird, sondern auch für den Weg, auf dem die Erkenntnisse erzielt wurden. Dieser Weg ist durch Selektionen bestimmt, die zusammen mit der Erkenntnis offenzulegen sind. Ohne Transparenz in diesem Sinne kann man unseres Erachtens gar nicht von Wissenschaft reden.

---

spricht man von *heavy weight ontologies*. Diese Ontologien werden mit mächtigen Modellierungswerkzeugen implementiert, die sowohl die Speicherung des ontologischen Schemas als auch den Aufbau der Inferenzmechanismen gestatten (z.B. Frame Logic).

5 Gewiss - es gibt ja auch die sog. common sense Ontologien (z.B. Word Net). Ihr Nutzen für den genannten Anwendungsbereich (automatisierbare Integration von Informationssystemen im Hinblick auf eine neue Leistung) ist allerdings nicht zu sehen.

Mit der Informatisierung des Wissens stellt sich das Problem der Diversifikation. An die Stelle gemeinsam geteilter Konzepte oder Modelle von Welt („Weltbilder“ im neuzeitlichen Verstande) treten fach- oder domänenspezifische Reichweiten und Geltungsbedingungen von Begriffen. Dies ist nun allerdings kein Nebeneffekt oder ein irgendwie behebbarer Mangel des modernen Wissenschaftsbetriebs. Diese Spezialisierung ist die Voraussetzung von Fortschritten im Fachwissen, sie ist die Grundlage für das erhöhte Auflöse- und Rekombinationsvermögen in den Wissenschaften, die immer kleinere Strukturen erforscht und immer komplexere Strukturen synthetisieren kann. Der neuzeitliche Wissenschaftsbegriff ist demzufolge schon längst überholt. Dieses Überholtsein betrifft vor allem die konstruktiven Bedingungen von Wahrheit. Wissenschaftliche Objektivität ist seit der Neuzeit das, was für ein Erkenntnissubjekt im Prinzip „transparent“ also „vor seinen Augen“ erzeugt, in seinen Konstitutionsbedingungen durchsichtig gemacht werden kann. Der hierfür benötigte gemeinsame Kontext - die Welt der wissenschaftlichen Tatsachen - lässt sich in diesem Sinne nicht mehr herstellen, allein schon deshalb, weil die fachwissenschaftlichen Begrifflichkeiten füreinander nicht mehr anschlussfähig sein können.

Architekturanstrengungen wie die des Semantic Web, bei denen Protokollschichten unsichtbar für den Nutzer die Auswahl der verfügbaren Information einschränken und mehr noch: mittels hochselektiver Filter Informationen aus unterschiedlichen Kontexten zu neuen Inhalten aufbereiten, vereinfachen diese Problematik keineswegs, sie *simplifizieren* sie. Sie schaffen damit selbst einen neuen und anderen Kontext von Wissen, den der Simplifikation. Anders gesagt: die Technologien zur Modellierung von Wissen schaffen neue Möglichkeitsräume, die unerkannt zusammen mit dem durch sie verfügbar gemachten Wissen „wuchern“.

Im technischen Sinne „transparente“ Verfahren zur automatischen Selektion von Informationen werden aus diesem Grunde in wirklich brauchbaren Systemen des Wissens- und Informationsmanagements immer durch eine Reihe flankierender Konzepte ergänzt werden müssen: etwa durch die Bereitstellung unterschiedlicher Verfahren, die unterschiedliche Ergebnismengen zurückliefern, durch Rückmeldungen des Systems, die auch die Grenzen der Genauigkeit oder Unzulänglichkeiten der Wissensbasis verdeutlichen, durch die Beteiligung von Nutzern bei der Strukturierung von Wissensbeständen und nicht zuletzt durch die Investition in „die Köpfe“, also durch Maßnahmen, die das schulen, was in der Philosophie „Urteilskraft“ genannt wird.

References

1. M. Minsky, Ein Rahmen für Wissensrepräsentation. In: Kognitionswissenschaft. Grundlagen, Probleme, Perspektiven. Hrsg. V. D. Münch. Frankfurt/Main 1992, S. 92-134.
2. Ludwig Wittgenstein: Philosophische Untersuchungen (aktuelle Ausgabe des 1953 als Nachlass herausgegebenen Werkes: Frankfurt/Main 2001), §§60-64..
3. Heide Balzert: Lehrbuch der Objektmodellierung, Heidelberg/Berlin 1999
4. Gruber, T. R. A Translation Approach to Portable Ontology Specifications, in. Knowledge Acquisition, 5(2):199-220, Academic Press, 1993.

Ontology: The Discipline and the Tool

Doug Mayhew[1,2], Dirk Siebert[1]

[1] Institute for Formal Ontology and Medical Information Science, University of Leipzig, 04107 Leipzig, Härtelstraße 16-18, Germany
doug.mayhew, dirk.siebert@ifomis.uni-leipzig.de
[2] Department of Philosophy, University at Buffalo, New York, USA

Abstract. The fields of philosophy and informatics entertain two somewhat different conceptions of ontology. Philosophical ontology is a branch of metaphysics dating back at least to the time of Plato and Aristotle. Ontology in informatics has its origins in the artificial intelligence research of the eighties and nineties. This means that the fields of philosophy and informatics entertain two somewhat different conceptions of ontology and the present paper discusses the relationship between these two conceptions. Differences and similarities are pointed out and variations in methodological approaches are also discussed. Efforts to combine the ontological methodologies and resources of the two fields are surveyed, and actual and potential benefits and drawbacks of such collaborations are examined.

Different Concepts

The fields of philosophy and informatics entertain two somewhat different conceptions of ontology, with correspondingly different notions of what ontology is for. First, we have the age-old conception of ontology as a philosophical discipline. Second, we have the relatively new conception of ontology as an information organization tool, a notion of ontology adapted from the philosophical conception by artificial intelligence researchers and then adopted by the applications-oriented field of informatics. (The philosophical conception was recognized in informatics as early as the late sixties in data modeling research [1].) The two conceptions have their characteristic differences, which show up primarily in what each field thinks ontology is for. The following is a discussion of these differences, but also of the similarities between the two disciplines and of the ways in which they might cooperate.

In philosophy, ontology is the study of what exists. Ontology is thus of a piece with metaphysics, the branch of philosophy that studies the nature of reality, and has been a prominent area of investigation since the beginnings of philosophical inquiry. The individual ontological theories advanced in philosophy are universal, descriptive classifications of the content and structure of reality as a whole. Aristotle provides a classic example in [2], where he describes the world using ten categories—viz., substance and nine kinds of attributes or accidents. Philosophical ontologies have traditionally been expressed informally, using natural language, but Leśniewski's [3] use of an artificial formal language to represent his formal theory of parts (mereology) inaugurated philosophy's use of artificial languages and formal logic in expressing ontologies. Examples of contemporary philosophical endeavors which involve the symbolical representation of ontologies can be found in [4] and [5].

Informatics uses ontology predominantly as an information organization tool. Many attribute the initial use of ontology in this way to artificial intelligence research on the facilitation of knowledge sharing and re-use among software agents. As described in [6], ontology in this field often entails the development of customized terminologies (with individually specified meanings for the terms) used to create customized descriptions or models of a particular domain of some actual or constructed reality. These are language dependent ontologies rather than universal theories. An example in the medical domain is SNOMED [7]. Ontologies in informatics are often expressed as logical theories (often using one or other flavor of description logic [8]), semantic networks, or, more recently, in a modeling language such as UML [9]. They might then be converted to programming language code to become part of a software application.

Whereas philosophical ontology has traditionally sought to provide a general ontology of reality as a whole, those working on ontologies in informatics develop domain-specific ontologies designed to meet particular information processing needs and requirements. This difference in scope points to yet another difference. The philosopher seeks knowledge of what exists more or less for the sake of knowledge itself. The informatician, in contrast, is interested in developing ontologies to serve more limited and practical purposes. There is normally very little theoretical work involved, for example, in developing an ontology of a company's product or service line; this is done simply to manage inventory and accounting records. The difference here is in one's motivation for doing ontology.

There are some other differences to note. First, as indicated above, philosophical ontologies are language-independent or universal, whereas many informatics ontologies are restricted to the specific languages in which they are formulated. Another difference turns on the fact that, while philosophy has traditionally been concerned with giving ontological accounts of the natural and continuous reality that everyone experiences, informatics is often working with the closed-world realities, for instance, of particular businesses, the specific products or services with which those businesses deal, the people and activities involved, and so forth—reflecting again their underlying, specifically pragmatic motivations.

Along with the differences, there are also similarities. In both fields, the general idea—leaving aside for the moment issues of domain, scope, degree of refinement, and method of expression—is that ontology provides a means to classify entities, processes, and the relations that hold between them. There is also the commonly—though not unanimously—held belief in both fields that ontology should strive for descriptive accuracy and cross-domain communicability.

Despite the many differences between the philosophical and informatics conceptions of ontology, a number of philosophers and informaticians are now working together in various ontology projects. Individuals from both fields are surveying the relevant literature and exchanging ideas in efforts to improve upon their respective uses of ontology. The benefits and drawbacks of such efforts are discussed in the last section of the paper.

Different Approaches

The conceptual differences discussed above are differences between fields. There are also different conceptual and methodological approaches to ontology within the two fields. An account of the different approaches in philosophical ontology can be found in [10], and a condensed account of philosophical and informatics approaches is provided in what follows.

Approaches in traditional philosophical ontology fall under two basic divisions, which we might call substance versus process and reductionist versus non-reductionist. Substance-based ontologies focus on substances or things as the essential constituents of reality. Process-based ontologies describe reality primarily in terms of processes, flux, or change. Reductionist approaches claim that reality is accurately described in terms of one basic type of constituent, usually either substance or process. It is rather difficult to find examples of pure substance or process reductionism, since almost all those who put forward a substance-based ontology also admit processes of one sort or another and vice versa. One finds an essentially substance-based, reductionist ontology in materialist metaphysical doctrines like that presented in [11]. An example of process-based reductionism can be found in [12]. Non-reductionist ontologies are concerned with providing an exhaustive ontological account of reality at all levels, from the micro- to the macroscopic, including categories of both substance and process. Aristotelian metaphysics is a good example, as is the ontological work in [13] and [14].

Within informatics there are also two basic divisions, the reference versus applications division and the logic-based versus non-logic-based division. Proponents of reference ontologies [5] advocate the creation of overarching, descriptively adequate ontologies accompanied by a rich formal representation. Applications ontology is focused on low-level ontologies designed to represent the taxonomical structure of specific domains, and proponents of applications ontology praise its advantages for the practical purposes of many information systems. Commerce, research, and information-based applications developed using ontological methods of organization range from simple yet extensive domain-specific terminologies to standards development projects involving software interoperability, information search and retrieval, automated inferencing, and natural language processing. A good example of applications ontology that handles a range of these tasks can be found in [15].

Examples of research being done in both reference and applications ontology can be found in [16, 17, 18]. Much of the commercial work done within ontology is applications ontology. Among reference and applications ontologists, one can find a further division between logic- and non-logic-based approaches. Logic-based approaches rely primarily on description logic to develop a model of some domain. Non-logic-based approaches, exemplified in network-based structures (e.g., connectionist systems and semantic networks), seek to develop models that more closely resemble the observable structure and workings of human cognition. An example of this approach is seen in [19].

Not every ontological effort in either field adheres strictly to one approach or the other. Hybrid approaches are common in philosophy and informatics, and such approaches often turn out to be useful in advancing both philosophical understanding and information systems performance. The non-reductionist approach in philosophy is itself a hybrid approach. One can find ontological projects in informatics and especially in AI that use logic in combination with non-logic-based approaches—e.g., in some of the latest software agent designs [20]. Then, of course, there are recent projects that combine approaches from the two fields of philosophy and informatics [16, 17, 18]. Hybrid approaches, as one might guess, tend to be richer in what they can express and how they express it, and thereby more complex and time consuming during the design phase.

Philosophy and Informatics Working Together

Though ontology has a long history in philosophy, it is probably safe to say that today it is known primarily from its association with artificial intelligence and informatics. This is based partly on the fact that there are many more people involved in doing research on ontology in informatics than in philosophy. There are, however, a few philosophers studying ontology as a discipline reaching beyond its philosophical origins. There are also informaticians exploring the philosophical origins of what they once thought of primarily as an information organization tool. Efforts to combine philosophical and informatics research in ontology are still in their infancy but growing steadily. Let us consider some of the actual as well as some of the potential benefits and drawbacks of philosophers and informaticians working together in ontology.

Benefits

Researchers on both sides stand to benefit, first of all, from simply being exposed to a different perspective. Combining perspectives is a good way to initiate the development of new ideas, and new ideas can turn into concrete improvements in the ways in which difficult problems are confronted. Following are some of the benefits, both mutual and one-sided, that could be reaped from the exchange of ideas and methods across the fields of philosophy and informatics.

Consider the difference between the grand-scale ontologies of philosophy and the domain-specific ontologies of informatics. Philosophers are trained to look at the big picture, to notice the content, structure, and relations of reality as a whole, and to develop a general ontological theory based on their investigations. Relations between the different aspects of reality can then be expressed in terms of the general ontological theory. This kind of training could be passed on to informaticians willing to exchange ideas with philosophers. Informaticians would learn how to develop a wider, yet cohesive view of reality into which their domain specific concerns could be integrated. One resulting potential benefit to informatics is a way to seriously improve upon systems and software interoperability and standards development.

While pure philosophical ontology is not pursued for the sake of practical ends, there are benefits that the philosopher might gain from studying informatics. Recall that informatics ontologies are often models of domains—e.g., law, commerce, and administration—where the entities are created entirely by the actions, physical and verbal, of human beings. These models can be generalized, and the result used to uncover prevalent features of our everyday lives. Philosophical efforts to account for that have surfaced only recently in works like [21]. Philosophical ontology should ultimately seek to give an account of every aspect of reality, and the examination of informatics ontologies could result in benefits to philosophers working toward this end by providing new families of examples and also new types of problems with which to grapple.

Attention to the conceptual differences and the different approaches observed between and within the two fields may reveal possibilities for hybrid approaches that can take advantage of the best features on either side. This is a good first step in developing rich and exhaustive ontologies. For example, an ontology or ontological method like that proposed in [22], that can account for different views or partitions of reality, would be useful in both reference and applications settings. If philosophers and informaticians continue to compare the results of their research efforts, then perhaps, to the benefit of both fields, a universal descriptive ontology that covers any kind of content at every level of granularity [22] can be developed.

There is a mutual benefit associated with the practice of formalizing ontologies. The primary reason for expressing an ontology in a formal language is to help achieve higher levels of clarity, rigor, and accuracy in ontological theories and models and to ensure that the content of the ontology is easily communicated to other people and perhaps more easily translated into programming code. Collaboration between philosophers and informaticians with a background in logic may assist in the creation of better formalizations.

A further benefit that can be realized from the combined efforts of philosophy and informatics is that each field will have more manpower working on its problems. The more there are working on a given problem, the better the chances it will get solved. Subjecting a body of work to more minds and different perspectives is also an effective way to expose previously hidden difficulties. Examples in which informatics has benefited in this respect can be found in [23] and [24].

The cooperation between fields has also created openings in knowledge engineering and consultancy that can be filled by philosophers with a background in ontology and logic. Companies like *Ontology Works*, *Cycorp*, *Kanisa*, and *Language and Computing* are among those that have put philosophers to work. Working in informatics has provided some philosophers with supplemental training that will undoubtedly be of value to them in the future.

These collaborative ontological efforts promise indirect benefits to those outside philosophy and informatics as well. If the collaboration produces improvements in information systems design and functionality, then there are subsequent benefits passed on to everyone who is served directly or indirectly by information systems. Clients, customers, and patients, for example, would certainly benefit from improved efficiency and accuracy in handling their needs. Businesses would improve their ability to better serve the client and thereby benefit financially from a stable or perhaps even growing clientele.

Drawbacks

In comparison to the actual and potential benefits attributable to the collaboration between philosophy and informatics, the drawbacks are few. It is difficult to conceive of any actual harm that could come from the two fields working together, but there are foreseeable difficulties and drawbacks.

An obvious difficulty arises with the attempt to assimilate concepts and approaches with which one is not entirely familiar. A lack of mutual knowledge and experience in the respective fields may result in the individuals talking past each other, rather than effectively communicating their ideas. The depth of professional knowledge possessed by a worker or researcher in one field may not easily transfer to someone in another, especially if the two fields are quite different. At the same time, years of training and practice in concepts and methods specific to a profession are hard to set aside, and we tend to think and communicate our ideas in terms of what we know, rather than in a context-free manner. Philosophy and informatics are sufficiently different for this difficulty to surface in collaborations between the two.

However, by providing the right environment for knowledge exchange this difficulty can be largely overcome. A learning environment and centers dedicated to philosophy and informatics research would help tremendously. (Such centers and special interest groups are being established in Buffalo, Leipzig, Rome, Trento, and Turin.) Along with sharing knowledge, supplemental training can help in overcoming this difficulty.

The difficulty with conceptual and methodological differences between the two fields points to a potential drawback concerning the time limits and predefined design guidelines imposed upon many informatics projects. Most philosophers do primarily theoretical work, and they are not under heavy time constraints. The work done in informatics is very often geared toward client-specific practical applications, and tasks must normally be completed within a relatively short time. An information system is often the backbone of operations; it is something a business, for instance, needs in order to do what it does. In other words, the people who need information systems cannot wait around while researchers figure out what the absolute best ontological design should be. The specific needs of the client must be taken into account, and the designers must develop the best system they can within the time limit set by the client.

There is a potential drawback for clients in that they must settle for what the systems designers can build for them in the time they are allotted. Put philosophers and informaticians to work on a project like this one, and we immediately see the drawbacks from their different perspectives. The philosopher wants to make sure that the ontology behind the system accurately reflects the world as it is and that any logical theory behind the design is sound and complete. The informatician may very well want the same thing, but realize that perfection must be sacrificed in the interest of finishing the job on time.

The immediate solution to this drawback is for the philosophically minded to learn how to work within predefined design guidelines and time limits. The ultimate solution is, again, to establish centers where basic research in philosophical and informatics ontology research can be carried out under conditions where time limits are dramatically relaxed. Research would be directed at developing better design guidelines in ways which could ultimately benefit everyone related to information systems.

Concluding Remarks

Collaboration between philosophy and informatics is a reality from which a number of benefits have been reaped and positive results produced. With the formation of more special interest groups, institutes, and centers dedicated to bringing philosophical and informatics ontology researchers together to learn from one another, it is reasonable to assume that more advances in both fields are on the way.

Acknowledgements

This work was supported by the Alexander von Humboldt Foundation under the auspices of its Wolfgang Paul Program.

References

1. Mealy, G. H.: Another Look at Data. In: AFIPS Conference Proceedings, Volume 31. Thompson Books, Washington, DC (1967) 525–534
2. Aristotle: Categories 1^b25. In: Barnes, J. (ed.): The Complete Works of Aristotle. Princeton University Press, Princeton (1984)
3. Lesniewski, S.: Foundations of the General Theory of Sets. In: Surma, S. J., Srzednicki, J. T. J., Barnett, D. I., Rickey, V. F. (eds.): Stanislaw Lesniewski: Collected Works, Volumes I and II. Kluwer, Dordrecht (1991)
4. Simons, P. M.: Parts: A Study in Ontology. Clarendon Press, Oxford (1987)
5. Smith, B.: Basic Formal Ontology. http://ontology.buffalo.edu/bfo (2002)
6. Guarino, N.: Formal Ontology and Information Systems. In: Guarino, N. (ed.), Formal Ontology in Information Systems: Proceedings of First International Conference, ACM Press, New York (1998) 3–15
7. Systematized Nomenclature of Medicine (SNOMED). http://www.snomed.org/
8. Baader, F., Calvanese, D., McGuinness, D., Nardi, D., Patel-Schneider, P., (eds.): The Description Logic Handbook: Theory, Implementation and Applications. Cambridge University Press, Cambridge, MA (2003)
9. Unified Modeling Language™ (UML™). http://www.uml.org/
10. Smith, B.: Ontology and Information Systems. http://ontology.buffalo.edu/smith/articles/ontologies.htm (2001)
11. Hobbes, T.: De Corpora. In: Molesworth, W. (ed.). The Collected English Works of Thomas Hobbes, Volume 1: London, Taylor and Francis Books Ltd. (1997)
12. Whitehead, A. N.: Science and the Modern World. Free Press, New York (1925)
13. Ingarden, R.: Time and Modes of Being. Michejda, H. R. (trans.), Charles C. Thomas, Springfield, Illinois (1964)
14. Chisholm, R. M.: A Realistic Theory of Categories: An Essay on Ontology. Cambridge University Press, Cambridge (1996)
15. Language and Computing. LinKSuite®. http://www.landcglobal.com/pages/linksuite.php
16. Institute for Formal Ontology and Medical Information Science (IFOMIS), Leipzig, Germany: Basic Formal Ontology (BFO) project. http://ontology.buffalo.edu/bfo/
17. Laboratory for Applied Ontology (LAO), Trento and Rome, Italy. http://www.loa-cnr.it/
18. Foundational Ontology. http://www.comp.leeds.ac.uk/brandon/ontology/

19. Unified Medical Language System® (UMLS®). http://www.nlm.nih.gov/research/umls/
20. Baral, C., Gelfond, M.: Reasoning Agents in Dynamic Domains. In Minker, J. (ed.): Logic-Based Artificial Intelligence, Kluwer, Dordrecht (2000) 257–279
21. Searle, J. R.: The Construction of Social Reality. Free Press, New York (1995)
22. Bittner, T., Smith, B.: A theory of granular partitions. In: Duckham, M., Goodchild, M. F., Worboys, M. F. (eds.): Foundations of Geographic Information Science, Taylor & Francis Books, London (2003) 117–150
23. Smith, B., Ceusters, W.: The First Industrial Philosophy: How Analytical Ontology Can Be Useful to Medical Informatics. Interdisciplinary Science Reviews, 28 (2003) 106–111
24. Kumar, A, Smith, B.: The Universal Medical Language System and the Gene Ontology: Some Critical Reflections. Proceedings of the 26th German Conference on Artificial Intelligence, Hamburg, 2003 (2003)

Using Philosophy to Improve the Coherence and Interoperability of Applications Ontologies: A Field Report on the Collaboration of IFOMIS and L&C

Jonathan Simon, Matthew Fielding, and Barry Smith, PhD.

Institute for Formal Ontology and Medical Information Science, University of Leipzig, Härtelstrasse 16, 04107 Leipzig, Germany
jonathan.simon@ifomis.uni-leipzig.de

Abstract. The collaboration of Language and Computing nv (L&C) and the Institute for Formal Ontology and Medical Information Science (IFOMIS) is guided by the hypothesis that quality constraints on ontologies for software application purposes closely parallel the constraints salient to the design of sound philosophical theories. The extent of this parallel has been poorly appreciated in the informatics community, and it turns out that importing the benefits of philosophical insight and methodology into application domains yields a variety of improvements. L&C's LinKBase® is one of the world's largest medical domain ontologies. Its current primary use pertains to natural language processing applications, but it also supports intelligent navigation through a range of structured medical and bioinformatics information resources, such as SNOMED-CT, Swiss-Prot, and the Gene Ontology (GO). In this report we discuss how and why philosophical methods improve both the internal coherence of LinKBase®, and its capacity to serve as a translation hub, improving the interoperability of the ontologies through which it navigates.

1 Introduction

We may understand an application ontology as a system of representations of elements of reality, structuring data according to some hierarchy of classes for the purpose of managing and manipulating that data, and supporting interoperability of various resources in automatic fashion. We may understand a philosophical ontology as a system of representations of elements of reality structured according to some hierarchy for the purposes of better understanding and relating those elements of reality to one another. These two forms of ontology can in principle support each other. The principal distinction is the demand, crucial in philosophical circles, that an ontology be maximally comprehensive. The philosopher strives for logical rigour, which means that she is not free to ignore irrelevant or rare counterexamples to her general schema. Such a demand is not present in many application ontologies, where the goal-driven context tends to encourage a view of such perfectionism as excessive and costly. Rather ad hoc algorithms are used which are designed to protect the system against counterexamples under given externally determined local conditions. More and more, however, researchers are coming to realize that this quick-fix methodology has not fulfilled its promise of bringing about interoperability between information resources, and that it has indeed hindered adaptability of preexisting systems to handle new applications, and to support new software.

As researchers from various fields – in the medical domain for example in fields such as natural language processing, clinical trials management, genetics research, anatomy representation and visualization – struggle with the same set of issues, they find themselves unknowingly appealing to the very same principles and methodologies that have driven philosophical research for thousands of years. The more global and flexible an application ontology strives to be, the more general are the data sets it must be prepared to manage, and the more it becomes possible to establish the beginning of an isomorphism between the data sets relevant to the domain in question, and the elements of reality represented by a maximally comprehensive philosophical ontology of that domain. Where the application ontologist evolves and tests his system in response to the cases actually presented by new data, the philosophical ontologist can evolve and test her system according to the methodology of the *Gedankenexperiment*, the practice of imagining possible scenarios which testify to the inadequacy of an existing representation. The thought experiments of ontologically-minded philosophers such as Aristotle, Brentano, Husserl, and Ingarden have in fact proven to be astonishingly prescient in anticipating the problems faced by application ontologies when new types of data need to be dealt with, and the responses these philosophers have suggested sometimes parallel the optimal revisions available to application ontologists in the relevant cases. [1],[2]

The hypothesis which drives the collaboration between the commercial enterprise Language and Computing (L&C) and the academic research group IFOMIS, the Institute for Formal Ontology and Medical Information Science, is that such parallels should be pushed as a matter of principle, and that the construction of an application ontology based on philosophical principles can yield considerable practical benefits. The procedure, broadly speaking, has been the callibration of LinKBase®, L&C's application ontology, in conformity with BFO, the philosophical ontology developed by IFOMIS. This callibration has already yielded positive benefits along two dimensions: 1) improving LinKBase®'s capacity to model certain types of data for the purposes of L&C's software applications, and 2) improving LinKBase®'s capacity to serve as a translation hub for ontologies like GO and SNOMED-CT by enabling the development of the mapping software MaDBoKs. In what follows we discuss these improvements, emphasizing the philosophical nature of the innovations which enable them, and drawing conclusions for the value of philosophical methodology in the advance of information systems. [3]

2 Methods

2.1 LinKBase® and BFO

LinKBase® is a biomedical domain ontology that has been designed to integrate terminologies and databases with applications designed for natural language processing and information retrieval. The ontology contains 543 different relations (linktypes), divided into different groups, including spatial, temporal and process-related link types. LinKBase® currently contains over 2,000,000 medical concepts organized in a graph with over 5,300,000 link type instantiations. Both concepts and links are language independent, but they are cross-referenced to about 3,000,000 terms in various languages. LinKBase® provides a central hub with fixed structured definitions into which external medical terminologies and databases may be embedded. This task turns out to be a complex endeavor, not least because the different terminologies or databases that are to be integrated are often internally and mutually inconsistent. Yet, as all these terminologies must essentially speak about the same reality, there is a common thread that runs through them and the LinKBase® methodology is based on the idea that it is possible to integrate them precisely by reference to those basic categorical distinctions that are common to them all.

Basic Formal Ontology (BFO) is a philosophically inspired top-level ontology which provides a coherent, unified understanding of these basic distinctions and which is currently being implemented as a top-level open source backbone ontology for LinKBase®. [4] BFO incoporates theories of continuants and occurrents, mereology, mereotopology, universals and particulars, biological classes (natural kinds) and instantiations, and of granular partitions, as well as respecting the more general demands on good ontology recognized by the philosophical community. [5] BFO is thus ideal as a framework for mapping external ontologies, terminologies, and databases onto LinKBase® in a way that is designed to provide for successful integration, and as a useful guide for the future algorithm development that will allow for cross-ontology navigation. The core of BFO is expressed as a simple is-a tree structure, with which is associated a more comprehensive first-order formalism, also available in a KIF representation. [6] In its logical manifestation, the richness of the BFO theory is exploited to guide changes and adaptations of the LinKBase® system. BFO is the result of collaboration among philosophers, linguists, computer scientists and physicians, and is currently being extended to a top-level formal ontology of biomedical categories such as function, site, system, anatomic structure, and so on.

2.2 First-Order Standardization

As ontologies and terminologies expand and are integrated together, it is natural that consistency will become increasingly difficult to maintain. One cause of this difficulty lies in the many ambiguities and inconsistencies that result from the lack of a standard unified framework for understanding those basic relations that structure our reality. The BFO formal ontology provides application ontologies with a set of standardized, first-order definitions for these ontological elements, definitions which can be exploited by reasoning applications, including applications designed for natural language understanding. By disambiguating the ontological structures underlying those informal definitions currently used, which characteristically fall below acceptable standards of formal precision, these formalizations can aid in the passage of domain knowledge between users and software agents, and thus improve coherence and adaptability in and between ontologies. [7]

The resultant standardization reflects an implementation of philosophical rigor along two dimensions. First, it establishes internal consistency on the basis of precise analyses of the concepts involved. Ontologies such as LinKBase® (as well as SNOMED and GO) are viewed as object languages with a certain "surface structure." They consist of systems of concepts joined together in binary relations such as is-a and part-of. For the most part however, these relations and concepts are given only in natural language and in a form that leads to various characteristic ambiguities. Thus, the project of defining a unique deep structure to which every such concept, relation, and axiom can be mapped requires sound conceptual analysis. The standardization effort gives us a methodology with which to identify and repair internal inconsistencies and ambiguities in LinKBase® and other ontologies.

The second dimension of rigor requires the use of the standard first-order logical language in which the concepts of BFO are defined and axiomatized. In this way the rigor of the BFO classification system is imported into an ontology from the outside. This importation is meta-ontological, in the sense that changes are not made directly within the external ontology itself; rather, their place in the BFO re-articulated domain ontology, in this case LinKBase®, is marked via an external mapping algorithm in a way that provides the degree of consistency required to navigate between different third-party ontologies.

The standardization on concepts, relations and axioms of LinKBase® runs as follows:

1) For every concept C, the definition consists in a mapping to a pair: < the class named by C, the extension of the class named by C >
2) For every relation R(X,Y), the definition consists in a mapping to a logical formula of the following form: For all x such that x is in the extension of the class named by 'X', there is a y such that y is an element in the extension of the class named by 'Y', and R*(x,y) (where R* is a relation in the formal language of BFO, for example part-of, defined as a relation between individuals, including those individuals which are instances of the classes with which we began).

Axioms, which are essentially instantiated relations, are defined by a mapping similar to the definition of relation presented above, differing only in that the variables are replaced by specific concepts within the ontology.

In the remainder of this essay we seek to accomplish two goals. We first examine ways in which the philosophical insights afforded by this standardization have allowed us to understand and resolve modelling errors within the LinKBase® ontology. We then discuss the way in which the BFO standardization has assisted in the effort of ontology integration in the biomedical domain.

3 Results

3.1 Resolving Ambiguities and Modelling Conflicts in LinKBase®

3.11 Objects and Processes in LinKBase®

In philosophical circles it is well understood that the universe accessible to our everyday cognition contains two types of entities that relate differently to time. There are on the one hand objects, such as tables, chairs, countries, and people. These entities are said to *endure* through time, which means that they do not have temporal parts but are rather wholly present at every moment in which they exist. On the other hand are processes like brain surgeries, heart attacks, lives. These are said to *perdure* through time, which means that they do have temporal parts, such as the first half of the surgery, the last phase of the heart attack, one's childhood. This distinction is not adequately made in existing application ontologies and taxonomies. In particular when the ultimate tribunal for those ontologies are natural language practices, it becomes very important to identify the ambiguity in terms like 'injury', 'dilation', and 'dislocation'. For each of these terms in fact corresponds to two distinct concepts. We speak both of an injury as a perdurant ('when did that injury occur?') and as an endurant ('that injury looks terrible'). Likewise with kinds of injuries, like dislocations: 'The dislocation of his shoulder occurred yesterday' vs. 'The doctor reduced the dislocation.' Indeed, in the medical domain it is commonplace for a sort of process and the state resulting from that process to share a name.

'Dilation' may stand for the process of dilation, i.e. of becoming broader: 'Once in place, a small balloon tip is inflated for a few seconds to *dilate* the artery.' Or, it may stand for the dilated, broadened structure: '*Dilation* of the posterior mitral ring was corrected.'

Here the philosophical distinction between endurants and perdurants allows us to maintain the separation of concepts which would otherwise be, and in standard medical terminologies often are, conflated. By implementing this distinction into the LinKBase® top level, we have been able to recognize these instances of homonymy when they appear. We thereby avoid a range of modeling errors that emerge in standard systems. [8],[9]

3.12 Absences in LinKBase®

It is a tenet of contemporary philosophy that absences are not entities in their own right, but rather, precisely, the absences of entities. Yet medical ontologies must represent natural medical language concepts like 'absence of bacteriuria (bacteria in the urine)', and 'sputum without blood'. Further, though less common, medical texts may feature reference to absences without a specified location of absence, because the location is determined by context.

The straighforward approach, and the approach that LinKBase® formerly used, violated the philosophical tenet mentioned above, and construed absences as special kinds of entities, called 'processes of absence'. With this approach, it was necessary to provide further specification of the processes in question. What kind of process is an absence? What is its duration? Who are its participants? How do we know when two descriptions of absences actually refer to the same entity?

Processes are perdurants, entities located in spacetime. They thus have boundaries, volumes, and locations ('the surgery took place in the operating room'). An adequate inference engine will know various things about such bounded objects: it will know, for example, that if the boundary of object x is different from the boundary of object y, then x cannot be the same object as y, and so on. In a natural language data extraction application, information about the boundary of an absence might be specified via a description like 'an absence *in the liver*.'

Philosophical scrutiny (one of whose functions is to test the adaptability of an ontology framework by demanding responses to creative counterexamples) tells us that the treatment of absences as processes is unstable. A reasoning engine attempting to handle and infer information about absences so construed runs the risk of deriving contradictions. This possibility arises when we need to establish whether differently described absences are identical. 'The book was absent from my apartment' and 'The book was absent from my bedroom' seem to refer to the same absence. However, as soon as we instruct our inference engine to consider the two absences here described as identical, we will encounter inconsistency. For the system will record both that the absence has as boundary: *my apartment*, and that it has as boundary: *my room*. But this is a contradiction, since of course x = y implies boundary_of(x) = boundary_of(y). How, then, should absences be treated in a more philosophically adequate framework?

Another tenet of philosophy is: distinguish the particular from the universal. When we say 'There is an absence of bacteria in the patient's urine' we clearly are not saying *of* the bacteria in the urine, that *it* is not there. Rather, we are saying *of the universal*: bacteria, that *it* has no *instances* in the patient's urine. Following this intuition, LinKBase®'s current modeling eliminates concepts of absence themselves. Rather, relations of absence (like: the absence of bacteria in the urine) are construed as relations between the relevant bacteria concept, and the urine concept, but here it is the universal bacteria that is involved: 'If x is the bacteria universal, and y is an instance of urine, then x has no instance located in y.' This technique allows us to make inferences very naturally that would be artificial and error prone on the basis of the absences-as-entities model. We no longer need to answer the question whether the absence of the book from my apartment is the same absence as that of the book from my room. Rather we may naturally infer that there is an absence of the book from my room, given that there is an absence of the book from my apartment. This will follow from our general knowledge of location and parthood.

Along with improving our reasoning power, this solution improves our representation structure, rendering applications involving absences more elegant and simple. The old representation of absences as processes blocked us from directly linking two entities where one entity is "absent in" the other entity. It forced, rather, the creation of a third concept: the process of "absence of entity" which related the two.

By representing absence in terms of universals and non-instantiation we avoid the need to create this third concept, and reduce the distance between the related concepts to one relation instead of two. (E.g. the concept "sputum without blood" can be represented with a direct link to the concept "blood", which will be interpreted formally as: 'The blood universal has no instance located in (the patient's) sputum'.) The distance between concepts, and between links on parent-child trees is relevant to many LinKBase® applications. [10]

3.2 How Philosophy Engenders Interoperability: GO and MaDBoKs

3.21 Objects and Processes within the Gene Ontology

The Gene Ontology (GO) is divided into three disjoint hierarchies: the *cellular component*, *biological processes*, and *molecular function* ontologies. [11] The first, equivalent to anatomy in the medical domain, is an ontology of endurants. It allows users to access the physical structure with which a gene or gene product is associated. A biological process, on the other hand, is defined in GO as 'a phenomenon marked by changes that lead to a particular result, mediated by one or more gene products.' This ontology is therefore a hierarchy of occurrents.

There are however some confusions over the role and nature of GO's molecular function hierarchy. While GO defines molecular function as 'the action characteristic of a gene product,' what biologists characteristically assert about functions makes it clear that functions do not occur, but rather endure; the function of a gene or gene product exists identically for as long as its bearer exists and it is present at all times, even if that function is never realized. Even mutant genes retain their function. Thus for example, "signal transducer activity" remains the function of the EPO_HUMAN protein even when the latter is incapable of performing the signal transduction process.

Molecular functions and biological processes are obviously closely related. The function "signal transducer activity" certainly *involves* performing "signal transduction" in some sense; yet in GO this relationship is undefined. The authors of GO have attempted to clarify the matter by stating, 'a biological process is accomplished via one or more ordered assemblies of molecular functions,' in order to suggest that the relation is one of agency. Here, functions *initiate* biological processes, but this would suggest that the one stands to the other in a relation of parthood, which GO on the other hand explicitly rules out. For GO's authors insist, correctly in our view, that parthood only holds between entities of the same hierarchy. So long as the associated relations continue to conflate the distinct categories of function and process within the ontology, however, architectural flaws in GO will continue to constrain the sorts of reasoning systems which can support. [12]

3.22 MaDBoKs: Philosophically Inspired Ontology Integration

The Mapping Databases onto Knowledge Systems tool (or MaDBoKS) is an extension of the LinkFactory® ontology management system that administers and generates mappings from external databases such as GO or Swiss-Prot onto LinKBase®. This mapping mediates the data contained in the external database in a manner that expands the hub ontology, leaving the structure of the foreign ontology untouched. The MaDBoKS system is designed in such a way that all implicit and explicit relationships between data from the different databases are mapped to the hub ontology. Administration of the mapping mediates the data contained in the different databases in such a way that it is associated with ontological information and the ontology is thereby virtually expanded with the data and relations from the external sources. In this manner we are able to navigate across problematic definitions and relations within an external database using the BFO standardization as translation mechanism.

We now discuss how this works in the case of GO. We first carefully investigated the top-layer categories of the three GO sub-domains that act as our gateway between the LinKBase® concepts and the remaining terms in GO. We identified the more general concepts of GO in LinKBase® and created new concepts in those cases where suitable equivalents were not already recognized. In this way we were able to relate GO's molecular function hierarchy to the two other GO hierarchies by integrating all three simultaneously into BFO.

In the case of the EPO_HUMAN protein example mentioned earlier, we established that by mirroring BFO defined structures, LinKBase® is able to appropriate this example and model the associated relations with an improved degree of clarity.

The connection between a protein and its function is captured in LinKBase® by a "has-function" relation, and the connection between a function and its corresponding processes is captured by the LinKBase® "realization" relation. The former reflects the relation between a substance and its function, and the latter that between a function and its expression or actualization. Clearly, this latter relation is skew to the whole/part relation, which is properly left exclusive to each hierarchy.

In this manner not only is GO consistently mapped to LinKBase®, but the expressiveness of GO itself has been expanded without any major alterations required in its core structure. [13]

4 Concluding Remarks

It is a tangled web we weave when we seek to create application ontologies without a basis in philosophically sound formal theories. The BFO formalism structuring LinKBase® yields clean data, improves the efficiency of LinKBase®'s own software applications, and supports the integration (and thereby the untangling) of data from different external data sources in a transparent way. It captures the intended semantics of the database terms, and filters out erroneous synonyms and other errors.

Support from the Wolfgang Paul Program of the Alexander von Humboldt Foundation is gratefully acknowledged.

References

1. Flett A, Dos Santos M, Ceusters W.: Some Ontology Engineering Procedures and their Supporting Technologies. EKAW2002 (2003)
2. Smith B., Rosse C.: The Role of Foundational Relations in the Alignment of Biomedical Ontologies. http://ontology.buffalo.edu/medo/isa.doc
3. Montayne F, Flanagan J.: Formal Ontology: The Foundation for Natural Language Processing http://www.landcglobal.com (2003)
4. Smith B.: Basic Formal Ontology. http://ontology.buffalo.edu/bfo (2002)
5. Smith, B.: Mereotopology: A Theory of Parts and Boundaries. Data & Knowledge Engineering (1996) 20:287-303
6. Anderson, W. and Grenon, P.: KIF Axiomatization of BFO. IFOMIS Report 01/04, http://www.ifomis.uni-leipzig.de (2004)
7. Smith B., Köhler J., Kumar A.: On the Application of Formal Principles to Life Science Data: A Case Study in the Gene Ontology. http://ontology.buffalo.edu/medo/Database_Integration.pdf
8. Ceusters W., Smith B., Kumar A., Dhaen C.: Mistakes in Medical Ontologies: Where Do They Come From and How Can They Be Detected? In Pisanelli D. ed., Ontologies in Medicine: Proceedings of the Workshop on Medical Ontologies, Rome, October 2003 Amsterdam, IOS Press, forthcoming
9. Kumar, A., Schulze-Kremer, S.: Revising the UMLS Semantic Network. Proceedings of MedInfo, September 7-11 2004. forthcoming
10. Van Geyt L, Martens P, Terzic B, Flanagan J.: Get More Out of Your Unstructured Medical Documents. http://www.landcglobal.com (2002)
11. GO (Gene Ontology General Documentation) http://www.geneontology.org/doc/GO.doc.html
12. Smith B., Williams J., Schulze-Kremer, S.: The Ontology of the Gene Ontology. Forthcoming in the Proceedings of AMIA 2003
13. Verschelde J.L., Dos Santos M, Deray T, Smith B, Ceusters W.: Ontology-Assisted Database Integration to Support Natural Langauge Processing and Biomedical Data-mining. Journal of Integrative Bioinformatics, forthcoming

A Formal Theory of Conceptual Modeling Universals

Giancarlo Guizzardi[1], Gerd Wagner[2], Marten van Sinderen[1]

[1] Centre for Telematics and Information Technology
University of Twente
Enschede, The Netherlands
{guizzard@cs.utwente.nl, sinderen@ctit.utwente.nl}

[2] Eindhoven University of Technology,
Faculty of Technology Management,
Eindhoven, The Netherlands
G.Wagner@tm.tue.nl
http://tmitwww.tm.tue.nl/staff/gwagner

Abstract. Conceptual Modeling is a discipline of great relevance to several areas in Computer Science. In a series of papers [1,2,3] we have been using the General Ontological Language (GOL) and its underlying upper level ontology, proposed in [4,5], to evaluate the ontological correctness of conceptual models and to develop guidelines for how the constructs of a modeling language (UML) should be used in conceptual modeling. In this paper, we focus on the modeling metaconcepts of *classifiers* and *objects* from an ontological point of view. We use a philosophically and psychologically well-founded theory of universals to propose a UML profile for Ontology Representation and Conceptual Modeling. The formal semantics of the proposed modeling elements is presented in a language of modal logics with quantification restricted to Sortal universals.

1 Introduction

Conceptual Modeling is regarded as a discipline whose importance spreads throughout several areas in the realm of Computer Science (e.g. Software Engineering, Information Systems Design, Domain Engineering, Database Design, Requirements Engineering, and Knowledge Engineering, among others). Its main objective is concerned with identifying, analyzing and describing the essential concepts and constraints of a universe of discourse with the help of a (diagrammatic) modeling language that is based on a small set of basic meta-concepts (forming a metamodel). Ontological modeling, on the other hand, is concerned with capturing the relevant entities of a domain in an ontology of that domain using an ontology specification language that is based on a small set of basic, domain-independent ontological categories (forming an upper level ontology). While conceptual modeling languages are evaluated on the basis of their successful use in the practice domain information modeling, ontology specification languages and their underlying upper level ontologies have to be rooted in principled philosophical theories about what kinds of things exist and what their basic relationships with each other are.

The Unified Modeling Language (UML) is a language initially proposed as a unification of several different visual notations and modeling techniques used for systems design [6]. UML is now a *de facto* standard for modeling computational systems and, recently, it has been proposed that the language should be also used as an Ontology Representation Language [7]. Moreover, in this paper the authors argue that although UML lacks a precise definition of its formal semantics, this difficulty shall be overcome with the current developments made by the precise UML community[1].

We believe, however, that defining constructs of a conceptual modeling only in terms of its mathematical semantics, although essential, it is not sufficient to make it a suitable ontology representation language. The position defended here is that, in order to model reality, a conceptual modeling language should be founded on formal upper-level ontologies. In other words, it should have both, formal and ontological semantics.

In a series of papers we have been employing the General Ontological Language (GOL) and its underlying upper level ontology, proposed in [4,5], to evaluate the ontological correctness of UML conceptual models and to develop guidelines that assign well-defined ontological semantics to UML

[1] http://www.cs.york.ac.uk/puml/

modeling constructs. In [1], we have discussed the meaning of the UML metaconcepts of *classes and objects*, *powertypes*, *association* and *part-whole* relations (*aggregation/composition*). The UML metaconcepts of *abstract classes* and *datatypes* are addressed in a companion paper [2]. The work presented here can be seen as a continuation of this work in which we focus on one aspect of the philosophical problem between universals and particulars (roughly, classifiers and object instances in UML terms).

Although classifier modeling constructs are fundamental in conceptual modeling (being present in all major conceptual modeling languages) there is still a deficiency of methodological support for helping the user of the language deciding how to model the elements of a given domain. In practice, a set of primitives is often used to model distinctions in different types of classifiers (Type, Role, State, Mixin, among others). However, the choice of how the elements that denote universal properties in a domain (viz. Person, Student, Red Thing, Physical Thing, Deceased Person, Customer) should be modeled is often made in ad hoc manner. Likewise, it is the judgment of what are the admissible relations among these modeling elements.

This paper proposes a philosophically and psychologically well-founded theory of universals (section 2). This theory is further used to generate a typology of UML classifiers together with a set of methodological guidelines that governs its use (section 3). Additionally, we provide a formal characterization of the types of universals proposed in section 2 in a language of modal logics with restricted quantification (section 4). Finally, section 6 elaborates on some conclusions and future work.

2 Towards a theory of classifier types for Conceptual Modeling: philosophical and psychological foundations

In [8], van Leeuwen shows an important syntactical difference in natural languages that reflects a semantical and ontological one, namely, the difference between common nouns (CNs) on one side and arbitrary general terms (adjectives, verbs, mass nouns, etc...) on the other. CNs have the singular feature that they can combine with determiners and serve as argument for predication in sentences such as: (i) *(exactly) five mice were in the kitchen last night; (ii) the mouse which has eaten the cheese, has been in eaten in turn by the cat.*

In other words, if we have the patterns *(exactly) five X...* and *the Y which is Z...*, only the substitution of X,Y,Z by CNs will produce sentences which are grammatical. To see that, we can try the substitution by the adjective *Red* in the sentence *(i): (exactly) five red were in the kitchen last night.* A request to 'count the red in this room' cannot receive a definite answer: Should a red shirt be counted as one or should the shirt, the two sleeves, and two pockets be counted separately so that we have five reds? The problem in this case is not that one would not know how to finish the counting but that one would not know how to start since arbitrarily many subparts of a red thing are still red.

The explanation for this feature unique of CNs lies on the function that determinates (demonstratives and quantifiers) play in noun phrases, which is to determine a certain range on individuals. Both reference and quantification requires that the thing (or things) which are referred or which form the domain of quantification are determinate individuals, i.e., their conditions for *individuation* and *identity* must be determinate. In other words, if it is not determinate how to count Xs or how to identify X that is the same as Y, the sentences in the patterns (i) and (ii) do not express determinate propositions, i.e. propositions with definite truth values.

The distinction between the grammatical categories of CNs and arbitrary general terms can be explained in terms of the ontological categories of Sortal and Characterizing universals [9], which are roughly their ontological counterparts. Whilst the latter supply only a principle of application for the individuals they collect, the former supply both a principle of application and a principle of identity. A principle of application is that in accordance with which we judge whether a general term applies to a particular (e.g. whether something is a Person, a Dog, a Chair or a Student). A principle of identity supports the judgment whether two particulars are the same, i.e., in which circumstances the identity relation holds.

In [10], Macnamara, investigates the role of sortal concepts in cognition and provides a comprehensive theory for explaining the process that a child undergoes when learning proper nouns and common nouns. He proposes the following example: suppose a little boy (Tom), which is about to learn the meaning of a proper name for his puppy. When presented to the word Spot", Tom has to decide what it refers to. One should notice that a demonstrative such as "that" will not be sufficient to determinate the bearer of the proper name? How to decide that "that" which changes all its perceptual properties is still Spot? In other words, which changes can Spot suffer and still be the same? As Macnamara (among others) shows, answers to these questions are only possible if Spot is taken to be a proper name for an individual, which is an instance of a Sortal universal. The principles of identity supplied by the Sortals are essential to judge the validity of all identity statements. For example, if for an instance of the sortal *Statue* loosing a piece will not alter the identity of the object, the same does not hold for an instance of *Lump of Clay*.

The statement that we can only make identity and quantification statements in relation to a Sortal amounts to one of the best-supported theories in the philosophy of language, namely, that the identity of an individual can only be traced in connection with a Sortal Universal, which provides a *principle of individuation* and *identity* to the particulars it collects [8,10,11,12]. The position advocated in this article affirms an equivalent stance for a theory of conceptual modeling. We defend that among the conceptual modeling counterparts of general terms (classifiers), only constructs that represent substance sortals can provide a principle of identity and individuation for its instances. As a consequence, the following principle can be postulated:

Postulate 1: Every Object in a conceptual model (CM) of the domain must be an instance of a CM-class representing a sortal.

As argued by Kripke [13], a proper name is a rigid designator, i.e. it refers to the same individual in all possible situations, factual or counterfactual. For instance, it refers to the individual Mick Jagger both now (when he is the lead singer of Rolling Stones and 60 years old) and in the past (when he was the boy Mike Philip living in Kent, England). Moreover, it refers to the same individual in counterfactual situations such as the one in which he decided to continue in the London School of Economics and has never pursued a musical career. We would like to say that the boy Mike Philip is identical with the man Mick Jagger that he latter became. However, as pointed out by Wiggins [14] and Perry [15], statements of identity only make sense if both referents are of the same type. Thus, we could not say that a certain Boy is the same Boy as a certain Man since the latter is not a Boy (and vice-versa). However, as Putnam put it, when a man x points to a boy in a picture and says "I am that boy", the pronoun "I" in question is typed not by Man but by a supertype of Man and Boy (namely, Person) which embraces x's entire existence [16]. A generalization of this idea amount to a thesis, proposed by Wiggins, named thesis D [14]: If an individual falls under two sortals in the course of its history there must be exactly one ultimate sortal of which both sortals are specializations. Griffin elaborates Wiggins' thesis D in terms of two correlated principles:

a) **The Restriction Principle:** if an individual falls under two distinct sortals F, F' in the course of its history then there is at least one sortal which F and F' are both specializations.
b) **The Uniqueness Principle:** if an individual falls under two distinct sortals F, F' in the course of its history then there is only *one ultimate sortal* which F and F' are both specializations. A sortal F is ultimate if there is no other sortal F' distinct from F which F specializes.

It is not the case that two incompatible principles of identity could apply to the same individual x, otherwise x would not be a viable entity (determinate particular) [8]. Imagine an individual x which is an instance of both Statue and Lump of clay. Now, the answer to the question whether loosing a piece will alter the identity of x is indeterminate since each of the two principles of identity that x obeys imply a different answer. As a consequence, we can say that if two sortals F and F' intersect (i.e. have common individuals in their extension), the principles of identity contained in them must be equivalent. Moreover, F and F' cannot supply a principle of identity for x, since both sortals apply to x only contingently and a principle of identity must be used to identify x all possible worlds. Therefore, there must be a sortal G that supplies the principle of identity carried by F and F'. This proves the restriction principle. The uniqueness of the ultimate sortal G can be argued as follows: (i) G is a sortal, since it supplies a principle of identity for all the things in its extension; (ii) if it restricts a sortal H then, since H cannot supply a incompatible principle of identity, H either: is equivalent to G (i.e. supply the same principle of identity) and therefore should be ultimate or does not supply a principle of identity for the particulars in its extension (see text on dispersive classifiers below). This proves the uniqueness principle. The unique ultimate sortal G that supplies the principle of identity for its instances is named a *substance sortal.*

As a consequence of the *uniqueness principle* we define a second postulate:

Postulate 2: An Object in a conceptual model of the domain cannot instantiate more than one CM-Class representing an ultimate Substance Sortal.

In the example above, the sortal Person is the *unique substance sortal* that defines the validity of the claim that Mick Jagger is the same as Mike Philip or, in other words, that Mike Philip persists through changes in height, weight, age, residence, etc… as the same individual. Person can only be the sortal that supports the proper name Mick Jagger in all possible situations because it applies necessarily to the individual referred by the proper name, i.e. instances of Person cannot cease to be so without ceasing to exist. As a consequence, the extension of a substance sortal is world invariant. This meta-property of classifiers is named *Modal Constancy* [12] or *rigidity* [17].

Sortals such as Boy and Adult Man in the example above, but also Student, Employee, Caterpillar and Butterfly, Philosopher, Writer, Alive and Deceased, which possibly apply to a continuant during a certain phase of its existence, are named phased-sortal in [14]. As a consequence of the *Restriction Principle* we have that for every phased-sortal PS that applies to a continuant, there is a substance sortal S of which PS is a specialization.

Contrary to substance sortals, phased-sortals apply to individuals contingently and, thus, do not enjoy modal constancy. For example, for an individual John instance of *Student*, we can easily imagine John moving in an out of the *Student* type, while being the same individual, i.e. without loosing his identity. Moreover, for every instance x of Student in a world w, there is another world w' in which x is not an instance of Student. This meta-property of classifiers is name *anti-rigid* in [17].

By considering how these different universals stand w.r.t rigidity another postulate can be derived:

Postulate 3: A CM-Class representing a rigid classifier cannot be a subclass a CM-Class representing an anti-rigid classifier

If PS is a phased-sortal and S is the substance sortal specialized by PS, there is a specialization condition ϕ such that x is a PS iff x is a S that satisfies ϕ [10]. A further clarification on the different types of specialization conditions allows us to distinguish between two different types of phased-sortals which are of great importance to the practice of conceptual modeling, namely, *phases* and *roles*.

Phases constitute possible stages in the history of a substance sortal. Examples are: (a) Alive and Deceased: as possible stages of a Person; (b) Catterpillar and Butterfly of a Lepidopteran; (c) Town and Metropolis of a City; (d) Boy, Male Teenager and Adult Male of a Male Person. *Classifiers representing phases constitute a partition of the substance sortal they specialize.* For example, if <Alive,Deceased> is a *phase-partition* of a sustance sortal Person then for every world w, every Person x is either and instance of Alive or of Deceased but not of both. Moreover, if x is an instance of Alive in world *w* then there is world *w'* such that x is not an instance of Alive in *w'*, which in this case, implies that x is an instance of Deceased in *w'*.

Contrary to phases, roles do not necessarily form a partition of substance sortals. Moreover, they differ from phases in terms of the specialization condition ϕ. For a phase P, ϕ represents a condition that depends solely on intrinsic properties of P. For instance, one might say that if Mick Jagger is a Living Person then he is a Person who has the property of being alive or, if Spot is a Puppy then it is a Dog which has the property of being less than a year old. For a role R, conversely, ϕ depends on extrinsic (relational) properties of R. For example, one might say that John is a Student then John is a Person who is enrolled in some educational institution or that, if Peter is a Customer then Peter is a Person who buys a Product y from a Supplier z. In other words, an entity plays a role in a certain context, demarcated by its relation with other entities.

Although Frege argued at length that "one cannot count without knowing what to count", in artificial logical languages inspired by him, natural language general terms such as CNs, adjectives and verbs are treated uniformly as predicates. For instance, if we want to represent the sentence "there are tall men", in the fregean approach of classical logic we would write $\exists x\ Man(x) \wedge Tall(x)$. This reading puts the count noun Man (which denotes a Sortal) on an equal logical footing with the predicate Tall. Moreover, in this formula, the variable x is interpreted into a "supposedly" universal kind Thing. So, the natural language reading of the formula should be "there are things which have the property of being a man and the property of being tall". Since, by postulate 1, all individuals must be instances of a substance sortal we must conclude that Thing is a unique universal ultimate sortal which is able to supply a principle of identity for all elements that we consider in our universe of discourse. Moreover, by postulate 2, this principle of identity must be unique. Can that be the case?

In [20], Hirsch argues that concepts such as Thing, (Entity, Element, among others) are *dispersive*, i.e. they cover many concepts with different principles of identity. For instance, in the extension of Thing we might encounter an individual x which is a cow and an individual y which is a watch. Since the principles of identity for Cows and Watches are not the same we conclude that Thing cannot supply a principle of identity for its instances. Otherwise, x and y would obey incompatible principles of identity and, thus, would not be determinate individuals. Therefore, as defended in [8,11,12,18], dispersive concepts do not denote sortals (despite the fact that they are considered CNs in natural languages) and *therefore cannot have direct instances.* More than that, since a principle of identity supplied by a substance sortal G is inherited by all classifiers that specialize G or, to put in another way, all subtypes of G carry the principle of identity supplied by G. Thus, all subclasses of a sortal are themselves sortals, ergo,

Postulate 4: A CM-Class representing a dispersive universal cannot be a subclass of a CM-Class representing a Sortal

3 An Ontologically well-founded UML profile for conceptual modeling

The Unified Modeling Language (UML) has built in extension mechanisms that allow one to modify the language elements to suite certain modeling needs. A coherent set of such extensions, defined accordingly to a specific purpose or domain, constitutes a *UML profile* [6].

A Stereotype is a lightweight extension mechanism that allows one to specialize UML modeling elements by defining additional constraints and sometimes a different graphical notation, so that they behave in some aspects as if they were instances of elements defined in new virtual metamodel. Stereotypes are also used to indicate difference in meaning or usage between modeling elements with a similar structure.

In [3], we have proposed a profile for UML to support the design of ontologically well-founded conceptual models according to the theory proposed in section 2. This profile (summarized in the table below) comprises of a set of stereotyped classes (specializations of the meta-construct class) that represents finer-grained distinctions between different types of substantial universals. Additionally, the profile incorporates a number of constraints that is applied to relations involving these stereotyped classes.

Stereotype	Description	Constraints
«kind» A	A kind represents a *substance sortal*, i.e. rigid, externally independent universals that supply a principle of identity for its instances. Examples could be instances of Natural Kinds (such as Person, Dog, Tree) and artifacts (Chair, Car, Television).	Every object in conceptual model using this profile must be an instance of a Kind, directly or indirectly (postulate 1). Moreover, it cannot be an instance of more than one ultimate Kind (postulate 2). A supertype of a kind cannot be a member of {« subkind », « phase », « role », « roleMixin »}
«subkind» A A	A subkind is a rigid, externally independent restriction of a kind which carries the principle of identity supplied by the kind. An example could be the subkind MalePerson of the kind Person. In general, the stereotype «subkind» can be omitted in conceptual models without loss of clarity.	A sypertype of a subkind cannot be a member of {« phase », « role », « roleMixin »}
«phase» A	It represents the phased-sortals *phase,* i.e. *anti-rigid* and *externally independent* universals defined as part of a partition of a kind. For instance, the partition {Catterpillar, Butterfly} of the kind Lepdopterum.	The phases $\{P_1 \ldots P_n\}$ that form a partition of a Kind K are defined in UML as a disjoint and complete generalization set. The kind K is always depicted as an abstract class.
«role» A	It represents a phased-sortal *role*, i.e. *anti-rigid* and *externally dependent* universal. For instance, the role student played by instance of the kind Person.	Roles and Phases are anti-rigid universals and cannot appear in a conceptual model as a superclass of a Kind (postulate 3). Moreover: Let X be a class stereotyped as « role » and r be an association representing X's restriction condition. Then, $\#X.r \geq 1$
«category» *A*	It represents a rigid and externally independent ***non-sortal***, a dispersive universal that aggregates essential properties which are common to different kinds. For example, the category RationalEntity as a generalization of Person and IntelligentAgent.	A category cannot have direct instances and must be depicted as an abstract class. A supertype of a category cannot be a member of {« kind », « subkind », « phase », « role », « roleMixin »}
«roleMixin» *A*	It represents an anti-rigid and externally dependent ***non-sortal,*** a dispersive universal that aggregates properties which are common to different roles. It includes formal roles such as *whole/part* and *initiatior/ responder*.	A role mixin cannot have direct instances and must be depicted as an abstract class. A supertype of a role mixin cannot be a member of {« kind », « subkind », « phase », « role »}. Let X be a class stereotyped as « roleMixin » and r be an association representing X's restriction condition. Then, $\#X.r \geq 1$

«mixin» A	The stereotype «mixin» represents properties which are essential to some of its instances and accidental to others (a meta-property named semi-rigidity in [17]). An example is the mixin *Seatable*, which represents a property that can be considered essential to the kinds Chair and Stool but accidental to Crate, Paper Box or Rock.	A mixin cannot have direct instances and must be depicted as an abstract class. A supertype of a mixin cannot be a member of {« kind », « subkind », « phase », « role », « roleMixin »}

4 A formal characterization of the proposed categories

In this section we provide a formal characterization of the notions discussed in section 2 by using a language L of quantified modal logics with identity. A model M in this language is a structure $<W,D,\delta>$ where W is a non-empty set of worlds, D is a non-empty domain of objects and δ is an interpretation function assigning values to the non-logical constants of the language. The domain D of quantification is that of *possibilia*, which includes all possible entities independent of their actual existence. Therefore we shall quantify over a constant domain in all possible worlds. Moreover, all worlds are equally accessible and therefore we omit the accessibility relation from the model structure. As a result we have a language that differs from the simplest language of quantified modal logic (QS5) [19] in two points. First, all quantification is restricted by special predicates called *sorts*. We adopt the following notation proposed in [12]:

(i) $(\forall S,x)\ A$ (ii) $(\exists S,x)\ A$

which can be read as *for every instance of S A holds* and *there is an instance of S such that A holds*, respectively.

In this article, (i) and (ii) are meta-linguistic abbreviations to the formulas $(\forall x\ S(x) \rightarrow A)$ and $(\exists x\ S(x) \wedge A)$, respectively, i.e., they conform to the Fregean analysis of restricted quantification. However, the primitive objects of quantification (elements of D) are continuants and as proposed in [8], the predicates used to restrict quantification represent the sortal universals that carry the principles identity, which are constitutive of the individuals that fall in their extension.

Second, individual constants of the language represent *proper names* of individuals (continuants) and, therefore, the interpretation function δ defined as

(iii) $\delta(c,w) \in D$, in which c is an individual constant
(iv) $\delta(S,w) \subseteq D$, in which S is a sort
(v) $\delta(P^n,w) \subseteq D^n$, in which P is a n-ary predicate

must obey the following constraint: for all $w,w' \in W$, $\delta(c,w) = \delta(c,w')$, i.e. the interpretation of an individual constant c (proper name) is world invariant. This amount to Kripke's thesis that proper names are rigid designators [13] and conforms to Montague's meaning postulate 1 (MP1) [8].

The quantification restricted in this way makes explicit what is only implicit in standard predicate logics. As previously discussed, suppose we want to state the following proposition : (a) *There are red tasty apples*. In classical predicate logic we would write down a *logical* formula such as (b) $\exists x\ (apple(x) \wedge tasty(x) \wedge red(x))$. In an ontological reading, (b) states that "there are things which are red, tasty and apple". The theory proposed section 2 rejects that we can conceptually grasp an individual under a general concept such as Thing or Entity or, what is almost the same, that a logic (or conceptual modeling language) should presupposed the notion of a *bare particular*. Moreover, it states that only a sortal (e.g. Apple) can carry a principle of identity for the individuals it collects, a property which is absent in attributions such as Red and Tasty. For this reason, a logical system when as used to represent a formalization of conceptual models, should not presupposed that the representations of natural general terms such as Apple, Tasty and Red stand in the same logical footing. For this reason, (a) should be represented as $(\exists Apple,x)\ (tasty(x) \wedge red(x))$ in which the sortal binding the variable x it is the one responsible for carrying its principle of identity.

Let F and G be two arbitrary universals such that F is specializes G. As a consequence we have that

1. $\Box\ (\forall F,x\ G(x))$

if G is a *rigid* universal then

2. $\Box(\forall G,x\ \Box G(x))$

or in other words, for all $w,w' \in W$ we have that $\delta(G,w) = \delta(G,w')$

For instance, Figure 1 depicts an example with the kind Person and its subkind Man. In this case we have the following instantiations of (1) and (2):

$\Box(\forall Man,x\ Person(x))$ $\Box(\forall Man,x\ \Box Man(x))$ $\Box(\forall Person,x\ \Box Person(x))$

In fact, in this example, the subkinds Man and Woman form a partition of the kind Person. In general, if $<U_1 \ldots U_n>$ is a partition of a universal U then we have that

3. $\Box(\forall U,x\ U_1(x) \oplus \ldots \oplus U_n(x))$

and, in this specific case, $\Box(\forall Person,x\ Man(x) \oplus Woman(x))$.

In the same figure 1, another partition is present, namely, the phase-partition Child, Adolescent, Adult of the kind Person. Phases are always defined as a partition and, thus, formula (3) always hold for a phase-partition $<K_1 \ldots K_n>$of a substance sortal S. Besides that, for all $K_a,K_b \in <K_1 \ldots K_n>$ such that $a \neq b$ we have that

4. $\Box(\forall K_a,x\ \Diamond K_b(x))$

in the example of figure 1,

$\Box(\forall Child,x\ \Diamond Adolescent(x))$ $\Box(\forall Child,x\ \Diamond Adult(x))$
$\Box(\forall Adolescent,x\ \Diamond Child(x))$ $\Box(\forall Adolescent,x\ \Diamond Adult(x))$
$\Box(\forall Adult,x\ \Diamond Child(x))$ $\Box(\forall Adult,x\ \Diamond Adolescent(x))$

Formula (4) implies

5. $\Box(\forall K_i,x\ \Diamond\neg K_i(x))$

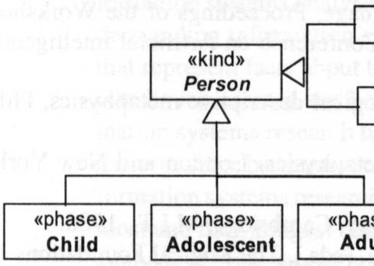
which is a more general statement of anti-rigidity and, hence, applies to all phased-sortals including roles. In figure 2, *Student* represents a role played by instances of the kind *Person*. As previously mentioned, roles differ from phases w.r.t. their specialization conditions. In figure 1, the association enrollment $\phi_{enrollment} \subseteq Student \times School$ represents a extrinsic property that must necessary apply to all instances of Student. In general, we can state the following: Let R be a role that specializes a sortal S (named its *allowed type*) and let ϕ be a relation representing the restriction condition for R, such that $\phi \subseteq R \times T$, where T represents a type on which R is *externally dependent* [17]. Then,

6. $\Box(\forall R,x\ \exists T,y\ \phi(x,y))$

in the case of figure 1

$\Box(\forall Student,x\ \Diamond\neg Student(x))$ $\Box(\forall Student,x\ \exists School,y\ \phi(x,y))$

Finally, we can show why the postulate 3 (section 2) must be reinforced in conceptual models. To see that is the case suppose there is a rigid classifier G which specializes an anti-rigid classifier F. Let *{a,b,c,d}* and *{a,b}* be the extension of F and G in world *w*, respectively. By (5), there is a world *w'* in which $a \in \delta(F,w)$ is not in $\delta(F,w')$ ($a \notin \delta(F,w')$). By (2), however, $\delta(G,w) = \delta(G,w')$ and, by (1), $\delta(G,w') \subseteq \delta(F,w')$, ergo, $a \in \delta(F,w')$ which is a contradiction. We have therefore shown that it is not the case that a rigid classifier could specialize an anti-rigid one.



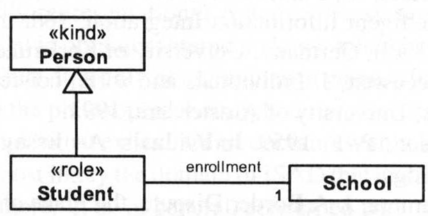


Figure 1 (left) – Example depicting a kind and two of its partitions: a subkind-partition and a phase-partition; Figure 2 – Example depicting a phased-sortal role, its allowed type and relational restriction condition.

5 Conclusions and Future Work

The development of a well-grounded, axiomatized upper level ontology is an important step towards the definition of real-world semantics for conceptual modeling diagrammatic languages. In this paper, we use a philosophically and psychologically well-founded theory of universals to address the problem of classifiers in conceptual modeling.

This theory is further used in the definition of a UML profile for Ontology Representation and Conceptual Modeling. The profile comprises of a set of stereotypes representing distinctions on types of classifiers proposed by the theory (e.g., Kind, Role, Phase, Category, Mixin) as well as a set of constraints on the possible relations to be established between these elements (representing the postulates of the theory).

A formalization of the theory is provided in a language of first-order modal logics with quantification restricted to Sortal universals. This formalization shall be extended in a future paper in which the difference between Sortals and arbitrary general terms will be emphasized. In particular, we intend to use *separated intentional properties* (in the spirit of Gupta's logic of Common Nouns [12]) to represent the intention of Sortal universals and to model the principles of identity and persistence supplied by them. This will enable us to formally address the notion of object state from an ontological point of view.

We believe that these results contribute to the task of defining ontological foundations and principled engineering tools for the discipline of conceptual modeling.

Acknowledgements Part of the work presented here has been developed during an extended visit of the first author to the ISTC/LOA (Trento, Italy) and in collaboration with the OntoMed research group (Leipzig, Germany). We would like to thank Nicola Guarino, Heinrich Herre, Claudio Masolo, Laure Vieu, Alessandro Oltramari, Stefano Borgo, Emanuele Botazzi, and Roberta Ferrario for fruitful discussions and for providing valuable input to the issues of this article.

References

1. Guizzardi, G., Herre, H., Wagner G.: Towards Ontological Foundations for UML Conceptual Models. In Proceedings of 1st International Conference on Ontologies Databases and Applications of Semantics (ODBASE 2002), Springer-Verlag, Berlin, Lecture Notes in Computer Science.
2. Guizzardi, G., Herre, H., Wagner G.: On the General Ontological Foundations of Conceptual Modeling. In Proceedings of 21th International Conference on Conceptual Modeling (ER 2002). Springer-Verlag, Berlin, Lecture Notes in Computer Science.
3. Guizzardi, G., Wagner G., van Sinderen, M. An Ontologically well-Founded Profile for UML Conceptual Models, In Proceedings of the 16th International Conference on Advanced Information Systems Engineering (CAiSE), Latvia, 2004. Springer-Verlag, Berlin, Lecture Notes in Computer Science.
4. Degen, W., Heller B., Herre H., Smith, B.: GOL: Towards an axiomatized upper level ontology. In Barry Smith and Nicola Guarino, editors, Proceedings of 2nd International Conference of Formal Ontologies and Information Systems (FOIS'01), Ogunquit, Maine, USA, October 2001. ACM Press.
5. Heller, B., Herre, H. 2003. Formal Ontology and Principles of GOL. Onto-Med Report No. 1/2003. Research Group Ontologies in Medicine, University of Leipzig.
6. Object Management Group, UML 2.0 Superstructure Specification, Doc.# ptc/03-08-02, Aug. 2003.
7. Cranefield, S., Purvis M.: UML as an ontology modelling language, Proceedings of the Workshop on Intelligent Information Integration, 16th International Joint Conference on Artificial Intelligence (IJCAI-99), Germany, University of Karlsruhe (1999) 46-53.
8. van Leeuwen, J. Individuals and sortal concepts : an essay in logical descriptive metaphysics, PhD Thesis, University of Amsterdam, 1991.
9. Strawson, P. F. 1959. Individuals. An Essay in Descriptive Metaphysics. London and New York: Routledge.
10. McNamara, J. A Border Dispute, the Place of Logic in Psychology, Cambridge, M.I.T., 1986.
11. McNamara, J., Logic and Cognition, in McNamara, J.; Reyes, G. (eds.), The Logical Foundations of Cognition (Vancouver Studies in Cognitive Science, Vol 4), October 1994.
12. Gupta, A.: 1980, The Logic of Common Nouns: an investigation in quantified modal logic, Yale University Press, New Haven.
13. Kripke, S., Naming and Necessity, Harvard University Press, 1982.
14. Wiggins, D., Sameness and Substance Renewed, Cambridge University Press, 2001.
15. Perry, J. *The same F.* Philosophical Review, 1970.
16. Putnam H., Logic and Psychology, in McNamara, J.; Reyes, G. (eds.), The Logical Foundations of Cognition (Vancouver Studies in Cognitive Science, Vol 4), October 1994.
17. Guarino, N.; Welty, C., A Formal Ontology of Properties, Proceedings of the 12th International Conference on Knowledge Engineering and Knowledge Management, 2000.
18. Hirsch, E. 1982. The Concept of Identity. Oxford University Press, New York, Oxford.
19. Fitting, M; Mendelsohn, R.L., First-Order Modal Logic (Synthese Library Studies in Epistemology Logic, Methodology, and philosophy Of Science Volume 277), Kluwer Academic Publishers, 1999.

On the Ontological Foundation of Modeling Grammars: A Critique

Boris Wyssusek

Technical University Berlin[1]

Abstract. Ontology and the concept of ontologies have attracted much attention in the context of research on modeling grammars for information systems analysis and design. Being rooted in philosophy, both ontology and the concept of ontologies bear their own history of philosophical debates which have quite often been ignored when applied in the field of information systems. In this contribution I claim that a more comprehensive discussion of well-known philosophical issues of ontology and ontologies will help us not only to understand the scope of their applicability in the context of modeling grammars. It will also provide us with insights about limitations as well as with directions for issues in need of further research. My argument is critical yet affirmative. It aims at the expansion of the scope of current debates and focuses especially on socio-philosophical aspects that need to be addressed in order to leverage the full potential of using ontology and ontologies for the provision of a theoretical foundation for modeling grammars.

1 Introduction

In the last two decades ontology and ontologies have gained considerable attention in the field of information systems research and practice, especially in the domain of information systems analysis and design (ISAD) (e.g., [2]; [7]; [11]; [18]; [31]; [33]). Understanding information systems as essentially representational systems, i.e., systems that represent facts about the "outside world," it is of major interest to know what there is to be represented and how to represent it. Thus, it is only of consequence when information systems research turns its attention to the philosophical discipline *ontology* that has ever since been concerned with "being" and "what exists." Of all domains within information systems research and practice it is most likely the domain of ISAD that has the most and the strongest ties to the world 'out there.' It is, in general, concerned with the analysis of "real-world" systems, the determination of changes that should occur in the "real-world" after the introduction or modification of an information system, and, eventually, the design of an information system.

An essential feature of ISAD is its use of models that, on the one hand, capture parts of the "real world" to be represented in the information systems, and, on the other hand, capture certain characteristics of the information system to be developed, e.g., its design. The acknowledged importance of modeling for ISAD finds its expression in the abundance of modeling grammars available and in the continuous efforts to improve

1. *Author's address:* Department of Computer Science, Technical University Berlin, Franklinstrasse 28/29, D–10587 Berlin, Germany. *E-mail:* wyssusek@cs.tu-berlin.de

these grammars as well as to develop new ones. Yet despite the abundance of modeling grammars, they hardly come with any theoretical foundation. And it is here where the interest of ISAD in ontology and ontologies arises (e.g., [26]; [27]; [28]).

The attraction of ontology is certainly due to its status as a well-established philosophical discipline, equipped with tradition, famous individuals, and a host of literature. As such we should welcome the possibility of information systems research opening up and drawing on the findings of this well-established discipline. But, as always, there are certain limitations we need to be aware of: Information systems researchers are seldom philosophers. Our drawing on a discipline we are not too familiar with is prone to the fallacies of gross misunderstandings, false analogies, and the like. Especially in philosophy, almost every term comes with its own history of debates; precise definitions are rare, and in order to fully grasp a notion we first need to know and to understand the theory (or theories) behind it. Nevertheless, there are good reasons why we should become engaged in philosophical theory. As COLLIER states, a "good part of the answer to the question 'why philosophy?' is that the alternative to philosophy is not no philosophy, but bad philosophy. The 'unphilosophical' person has an unconscious philosophy, which they apply in their practice—whether of science or politics or daily life" ([8], p. 17). Since the meaning of ontology and ontologies is bound to the respective philosophical theories that are being used as horizons of interpretation, we should not only be aware of their immediate consequences for the understanding of ontology and ontologies but should also be aware of more distant consequences that derive from the use of the respective theory (or theories).

A quite common understanding of ontology and ontologies in the ISAD literature is based on the ontological theory of BUNGE ([3]; [4]; [5]). As such it is grounded in a rather materialist-realist philosophical position that hardly finds any support in contemporary philosophy and social sciences. Consequently, when we consistently follow the ideas of BUNGE, it is impossible to connect the ontological foundation of modeling grammars with most of the contemporary literature dealing with the social world that ultimately provides us with everything to be represented in information systems. Hence, the uncritical adoption of BUNGE's ontological theory not only restricts our understanding of the world but also narrows our view to such an extent that we become unable to recognize the limitations of this theory as well as the negative ethical consequences of its application.

It is not my intention to dogmatically counter BUNGE's theory with yet another. Rather I ask the reader to engage in critical reflection when s/he is about adopt any ontological theory as a foundation for modeling grammars. In HABERMAS' ([16]) sense, the ontological foundation of modeling grammars should not solely be guided by an instrumental and/or a hermeneutical cognitive interest but also by an emancipatory interest that eventually helps to overcome self-inflicted cognitive constraints.

2 On ontology, ontologies and modeling grammars

During the last two decades the words "ontology" and "ontologies" have become quite popular in the literature dealing with the theoretical foundation of modeling grammars. But already a superficial reading reveals that there are many different understandings of those words. The fact that the spell-checker of my word-processor rejects the plural of ontology seems to be a rather far-fetched motivation for the closer examination of the words "ontology" and "ontologies", yet it points directly to a fundamental issue.

Even if early Greek philosophers were already concerned with ontological problems, it was only in the 17th century that the word "ontology" was introduced to denote a branch of philosophy (contemplative sciences) ([12], p. 16). Refraining from confronting the reader with the etymology of the word "ontology," I restrain myself to the exploration of a distinction which I believe to be most useful for the understanding of what ontology is all about. The following encyclopedic definition will serve as starting point for this purpose: "The word 'ontology' is used to refer to philosophical investigation of existence, or being. Such investigation may be directed towards the concept of being, asking what 'being' means, or what it is for something to exist; it may also (or instead) be concerned with the question 'what exists?', or 'what general sorts of things are there?'" ([9]).

I appreciate this definition but I regard it as flawed in one major respect: it might lead the reader to the conclusion that both questions—"what does it mean for something to exist?" and "what exists?"—can be answered independently. But it is trivial to realize that we cannot answer the question "what exists?" without having answered the question "what does it mean for something to exist?" first. With HEIDEGGER's words: "Basically, all ontology, no matter how rich and firmly compacted a system of categories it has at its disposal, remains blind and perverted from its own aim, if it has not first adequately clarified the meaning of Being, and conceived this clarification as its fundamental task" ([17], p. 31). Another important point was made by KANT who criticized the very idea of ontology. According to him, ontology "presumptuously claims to supply, in systematic doctrinal form, synthetic a priori knowledge of things in general" ([20], B303). In his argument he opposes realism to idealism: realism means that we perceive objects whose existence and nature are independent of our perceptions, whereas idealism means that they are dependent on our perception. Not satisfied with both positions, he argues: "Thoughts without content are empty, intuitions without concepts are blind. It is, therefore, just as necessary to make our concepts sensible, that is, to add the object to them in intuition, as to make our intuitions intelligible, that is, to bring them under concepts. These two powers or capacities cannot exchange their functions" ([20], B75). KANT reversed the classical view of epistemology. Instead of understanding knowledge as conforming to objects, we have to understand the objects as conforming to the conditions of the possibility of our knowing. Thus, human knowledge is limited to appearances; we are not able to know of the "things-in-themselves"—ontology cannot tell us anything about "things-in-themselves." KANT brought to our attention that all ontology is epistemic bound. Hence, ontology without epistemology is without any merit. This idea has been fully developed by HEIDEGGER. In response to KANT, to whom the "scandal of philosophy" was that no proof has yet been given of the "existence of things outside of us" ([20], Bxl), HEIDEGGER argues that the scandal is "not that this proof has yet to be given, but that such proofs are expected and attempted again and again" ([17], p. 249). Following HEIDEGGER, the question for the nature of reality of the external world poses a pseudo-problem.

In classical philosophy ontology was understood as a science, or as a discipline within or next to metaphysics. As such, there was no plural of the word "ontology." We do not speak about "biologies" just because different scientists have different understandings of biology. I do not know when and who it was who first used the word ontology in its plural form, but the idea of multiple ontologies has always been connected with philosophical issues of language. In his studies on language, HUMBOLDT came to the conclusion that people of different languages construct their world differently ([19]). NIETZSCHE argued that our entire understanding and knowledge of the world is bound to our language and—denying any objective meaning, or any actual reference of language to 'the world'—that it is of essentially metaphorical nature ([22]). The biologist UEXKÜLL argued that different species live in different worlds since their modes of cognition are structured differently ([25]). CASSIRER draws on the findings of UEXKÜLL and develops the idea that the world we consciously live in is essentially a symbolic world. The structure of this world does not so much depend on the 'outside' world rather than on the ways humans interact socially by means of symbolization ([6]). SAPIR and WHORF, furthering the idea of linguistic relativism, also contributed substantially to our understanding of the role of language in the construction of our world ([24]; [32]). It seems warranted to say that almost the entire philosophy of the 20th century was more or less concerned with the relationship between language and cognition.

With this history in mind it does not come as a surprise when some researchers, who are concerned with the linguistic representation of knowledge about the world, call their constructs "ontologies" (e.g., [14]; [15]). But what has often been ignored is that these researchers are well aware of the differences between ontologies they construct and ontology in its philosophical sense: "The word 'ontology' seems to generate a lot of controversy in discussions about AI. It has a long history in philosophy, in which it refers to the subject of existence. [...] In the context of knowledge sharing, I use the term ontology to mean a specification of a conceptualization. That is, an ontology is a description (like a formal specification of a program) of the concepts and relationships that can exist for an agent or a community of agents. This definition is consistent with the usage of ontology as set-of-concept-definitions, but more general. And it is certainly a different sense of the word than its use in philosophy. [...] [A]n ontology is a specification used for making ontological commitments. [...] Practically, an ontological commitment is an agreement to use a vocabulary [...] in a way that is consistent (but not complete) with respect to the theory specified by an ontology" ([14]). Hence, ontologies are linguistic conventions that do not tell us anything about the 'outside' world.

Taking linguistic relativity seriously, the play of language structures our (symbolic) world. We use language to objectify our experiences, thereby making it possible to communicate these experiences even if they lie in the past. Objectification also enables us to project potential future experiences and to communicate these projections ([1]). If language is so closely knit to our experiences it comes as quite natural to understand language as a means of representation of our experiences. Yet, there is another feature of language we have to take into consideration: The meaning of linguistic expressions is not fixed—symbols are multivalent. In short, we use the same expression to express different meanings. Quite often objectification has been confused with objective meaning of linguistic expression. And, if GRUBER understands ontology as a description or a specification of conceptualizations available to an (artificial) agent, then we must be aware of the fact that an artificial agent does not conceptualize. Artificial agents only command over a linguistic structure, not over conceptualizations. Concepts and the meaning of linguistic structures always remain in the realm of the human mind.

If we are about to develop an ontological foundation for modeling grammars we have to differentiate between the two distinct meanings of ontology depicted above. Referring to the BUNGE-WAND-WEBER (BWW) ontology, this distinction is quite obvious. On the one hand we are familiar with all the constructs that are descriptions or specifications of conceptualizations (e.g., [29], p. 64). Thus, these constructs are parts of an ontology—according to GRUBER's definition. On the other hand, we are familiar with the claim that these constructs are the 'things' that make up the world ([28], pp. 213 ff.). This claim belongs to the realm of philosophical ontology. And it is this claim that is highly questionable since it is based on a rather materialist-realist philosophical position that not many would subscribe to these days (for some critical self-reflection see, e.g., [27]; [30]; [31], pp. 174 ff.). According to this position, an ontological model is regarded as a representation (mapping) of the "true reality," or, given perceptions. This representational notion of "model" presupposes a direct relationship between the model (the representation) and the model source (the original). A model is "good" or "true" if it corresponds with reality—the essence of the correspondence theory of truth. Accordingly, for the development of an ontological foundation for modeling grammars it is decisive to 'find' the true objects and relationships in the "real world." But as far as I know, nobody knows how to do this. The "scandal" is that there are still people around who try to find the true objects and relationships in the "real world."

Summing up, the effort geared towards the development of an ontological foundation for modeling grammars will only be fruitful if we dismiss the idea of an ontological foundation in the classical philosophical (or metaphysical) sense. Our symbolic world, structured by means of language and symbolic social interaction, is actually a plurality of worlds. Therefore, I recommend that the followers of BUNGE, WAND, and WEBER do not adhere to the ontological, epistemological, methodological, and anthropological position held by BUNGE. Rather, it seems to be more promising to understand "ontological foundation" in the sense of GRUBER. The constructs provided by the BWW ontology do not need to be grounded in some metaphysical theory. They might serve us well if we understand them as descriptions or specifications of conceptualizations. If we commit ourselves to this ontology we commit ourselves to a grammar that might or might not be useful when we speak about the world. We will know of its usefulness if we try to express our conceptualizations by means of this ontology. If it does not serve our purpose we do not have to change our worldview—changing the grammar will suffice. The notion of linguistic relativism will help us to understand why people understand an ontology differently, or, in other words, why they attribute different meanings to one and the same ontology. And, it will also direct us to a question that especially needs our attention: How do we develop conceptualizations that are at least compatible in such a way that we are able to communicate by means of an ontology we have subscribed to?

3 Summary and conclusion

The persistence of the software crisis—budget overruns, exceeded time frames, not meeting user's requirements, and the total failure of information systems development projects—provides a host of motivations for the reconsideration of the state-of-the-art in information systems development. In order to overcome these highly undesirable results it is widely believed that the development of more rigorous theoretical foundations for modeling grammars is the key to success. One approach that has gained considerable attention during the last two decades is the development of ontological foundations for modeling grammars for ISAD.

Quite in contrast to the general lack of interest in philosophical issues of information systems, ontology has become an inspiring source for information systems research and practice. Understanding information systems as essentially representational systems, i.e. as systems that represent knowledge about certain domains, ontology is largely being used for the identification and/or the definition of that what constitutes these domains. Thus, information systems research is more interested in the ontological question "what exists?" rather than in the question "what does it mean to exist?"—a preference and a negligence with considerable consequences.

It was the linguistic turn in analytical philosophy that brought the language dependency of all knowledge to our attention. Among others, HUMBOLDT, CASSIRER, SAPIR, WHORF and the mature WITTGENSTEIN argue that our access to the world is bound to language and that there is no way to transcend our knowledge beyond the means provided by language: "The limits of my language mean the limits of my world" ([34], § 5.6). With the linguistic turn we have dismissed classical ontology. Ontology is bound to language, and the question for existence has been turned into a question of ontological commitment: "To be is to be a value of a variable within a given theory" ([23]). This understanding of ontology is accompanied by the acceptance of the existence of multiple ontologies and the reduction of philosophical ontology to the study of language. The problems associated with everyday language were the starting point for the development of formal languages which in turn provide the basis for the development of ontologies. But formal languages (and ontologies) do not help to overcome the problems of subjectivity and linguistic relativism. The constructs of formal languages have per se no meaning in our "life-world" (Lebenswelt). Formal semantics do not tell us anything about 'our' world, that is, a world of informal languages. Their meaning is derived from symbolic social interaction and is always relative to communities of practice. It is only in such communities that objectifications by means of language develop a stable yet not fixed meaning that enables the members of the respective community to communicate efficiently and effectively. If we intend to say something about the world by means of ontologies, we have to develop a common language practice that eventually will lead to the desired stable meanings. So far, this problem has hardly been addressed in the literature dealing with the ontological foundation of modeling grammars for ISAD.

In this contribution I have argued that the restricted scope of discourses and debates on the ontological foundation of modeling grammars for information systems analysis and design is partly due to the negligence of well-documented debates within the information systems research community. If we focus entirely on formal aspects of the ontological foundation of modeling grammars we are prone to fall victim to the error of the third kind, i.e., finding the right answers to the wrong questions. For example: Why bother with the evaluation of modeling grammars?—If we are convinced of an ontology then why not using this very ontology as modeling grammar?

Current research on the evaluation of modeling grammars in the context of ISAD has provided proof of the usefulness of ontologies for certain purposes (e.g., [13]; [21]; [10]). But we need to keep in mind that such a proof only proves that one axiomatic system conforms to another axiomatic system, or, that it does not. We have no proof that the axiomatic reference system—the ontology—is suitable for expressing something about the world. And, we still have to prove that the analytical approach toward the development of the axiomatic reference system exemplified by the BWW ontology is superior to, e.g., phenomenological or hermeneutical approaches.

References

[1] Berger, P.L. & Luckmann, T. (1991). *The social construction of reality.* London: Penguin

[2] Boland, R.J. jr. (1982). The Process of System Design: A Phenomenological Approach. In: M. Ginzberg, M. & C.A. Ross (eds.), *Proceedings of the 3rd International Conference on Information Systems* (pp. 31–45)

[3] Bunge, M. (1977). *Treatise on Basic Philosophy: Volume 3: Ontology I: The Furniture of the World.* Boston: Reidel

[4] Bunge, M. (1979). *Treatise on Basic Philosophy: Volume 3: Ontology II: A World of Systems.* Boston: Reidel

[5] Bunge, M. (1993). Realism and Antirealism in Social Sciences, *Theory and Decision*, 35:207–235

[6] Cassirer, E. (1967): *An essay on man. An introduction to a philosophy of human culture.* (eighteenth edition), New Haven: Yale University Press

[7] Checkland, P. (1981). *Systems Thinking, Systems Practice.* Chichester: John Wiley & Sons

[8] Collier, A. (1994). *Critical realism: an introduction to the philosophy of Roy Bhaskar.* London: Verso

[9] Craig, E. (1998): Ontology, In E. Craig (ed.), *Routledge Encyclopedia of Philosophy. Volume 7, Nihilism – Quantum mechanics, interpretation of.* London: Routledge

[10] Fettke, P. & Loos, P. (2003). Ontological Evaluation of Reference Models using the Bunge-Wand-Weber Model. In *Proceedings of the Ninth Americas Conference on Information Systems.* (pp. 2944–2955)

[11] Floyd, C. (1992). Software Development as Reality Construction, In C. Floyd, H. Züllinghoven, R. Budde & R. Keil-Slawik (eds.), Software *Development and Reality Construction* (pp. 86–100), Berlin: Springer

[12] Goclenius, R. (1613). *Lexicon philosophicum, quo tanquam clave philosophiæ fores aperiuntur.* Frankfurt/Main

[13] Green, P. & Rosemann, M. (1999). An Ontological Analysis of Integrated Process Modelling, In M. Jarke & A. Oberweis (eds.), *Proceedings of the 11th International Conference on Advanced Information Systems Engineering* (pp. 225–240). Berlin: Springer

[14] Gruber, T.R. (1993): What is an Ontology? <http://www.ksl.stanford.edu/kst/what-is-anontology.html>

[15] Guarino, N. (1995). Formal Ontology, Conceptual Analysis and Knowledge Representation, *International Journal of Human-Computer Studies*, 43(5/6):625–640

[16] Habermas, J. (1972). *Knowledge and Human Interests*. Boston: Beacon Press

[17] Heidegger, M. (1962). *Being and Time*, New York: Harper and Row

[18] Hirschheim, R., Klein, H.K., & Lyytinen, K. (1995). *Information systems development and data modeling: Conceptual and Philosophical Foundations*. Cambridge: Cambridge University Press

[19] Humboldt, W. von (1963). *Über die Verschiedenheit des menschlichen Sprachbaues und ihren Einfluß auf die geistige Entwicklung des Menschengeschlechts (1830 bis 1835)*, edited by A. Flitner & K. Giel, Darmstadt: Bibliographische Institut

[20] Kant, I. (1787). *Critique of Pure Reason* (second edition)

[21] Milton, S., Kazmierczak , E., Thomas , L. (2000). Ontological Foundations of Data Modeling in Information Systems. In *Proceedings of the Americas Conference on Information Systems 2000* (pp. 1537–1543)

[22] Nietzsche, F. (1873): *On Truth and Lie in an Extra-Moral Sense.*

[23] Quine, W.V.O. (1961): From a Logical Point of View: 9 Logico-Philosophical Essays. (second edition), Cambridge: Harvard University Press

[24] Sapir, E. (1949). *Selected writings of Edward Sapir in language, culture and personality.* Berkeley: University of California Press

[25] Uexküll, J. von (1934). *Streifzüge durch die Umwelten von Tieren und Menschen. Ein Bilderbuch unsichtbarer Welten*. Berlin: Springer

[26] Wand, Y. (1996). Ontology as a Foundation for Meta-Modelling and Method Engineering, *Information and Software Technology*, 38:281–287

[27] Wand, Y., Monarchi, D.E., Parsons, J. & Woo, C. (1995). Theoretical foundations for conceptual modeling in Information Systems Development, *Decision Support Systems*, 15:285–304

[28] Wand, Y., & Weber, R. (1988). An Ontological Analysis of some fundamental Information System Concepts, In J.I. DeGross & M.H. Olson (eds.), *Proceedings of the Ninth International Conference on Information Systems* (pp. 213–224)

[29] Wand, Y. & Weber, R. (1990a). Toward a Theory of the Deep Structure of Information Systems, In J.I. DeGross, M. Alavi & H. Oppelland (eds.), *Proceedings of the Eleventh International Conference on Information Systems* (pp. 61–71)

[30] Weber, R. (1997). The Link between Data Modeling Approaches and Philosophical Assumptions: A Critique, In *Proceedings of the Americas Conference on Information Systems 1997*

[31] Weber, R. (1997a). *Ontological Foundations of Information Systems*. Coopers & Lybrand Research Methodology Monograph No. 4, Melbourne: Coopers & Lybrand

[32] Whorf, B.L. (1956): Language, Thought, and Reality. Selected Writings of Benjamin Lee Whorf. Cambridge: MIT Press

[33] Winograd, T. & Flores, C.F. (1986). *Understanding Computers and Cognition. A New Foundation for Design*. Norwood: Ablex

[34] Wittgenstein, L. (1922): *Tractatus logico-philosophicus*. London: Routledge and Kegan Paul

Philosophical Issues in Computer Science

Ralph Kopperman[1], Steve Matthews[2], and Homeira Pajoohesh[3]

[1] Department of Mathematics, The City College of New York,
138th St & Convent Avenue, New York, NY 10031, USA.
rdkcc@att.net
[2] Department of Computer Science,
University of Warwick, Coventry CV4 7AL, UK.
Steve.Matthews@warwick.ac.uk
[3] CEOL, Unit 2200, Cork Airport Business Park,
Kinsale Road, Co. Cork, Ireland.
homeiraaa@yahoo.com

Abstract. The traditional overlap between computer science and philosophy centres upon the issue of in what sense a computer may be said to *think*. A lesser known issue of potential common interest is the ontology of the semantics computer science assigns to programming languages such as Pascal or Java. We argue that the usual purely mathematical approach of denotational semantics necessarily excludes important semantic notions such as *program reliability*. By studying the ontology of determinism versus non determinism we deduce the presence of *vagueness* in programming language semantics, this then being compared to the *Sorites paradox*. Parallels are drawn between Williamson's treatment of the paradox using *ignorance* and generalised metric topology.

Introduction

Computer science has well established itself as an academic discipline throughout universities of the world. Often a separate department will exist specifically for the large number of undergraduates who wish to major in the subject they loosely term 'computers'. The power which computers bring both to the individual and society to change the world, for better or for worse, is self evident. What is less clear is the extent to which the use of computers changes our comprehension of the world. While internet communication makes the world appear smaller it does not mean we have created a new world as such. However, popular culture can clearly visualise the existence of inorganic life forms such as Commander Data in Star Trek. As there is no conceptual problem for *android* life forms can we in the academic discipline of computer science not provide substance to the notion of a computer *existing*? Usually this question is reduced to, can a computer *think*? If so, it is considered to exist, otherwise not. But this begs the question, what is *intelligence*? While being an interesting question, it does not address matters of computer existence which, while successfully employed in computer science, bear no obvious relation to intelligence.

Due to Gödel's incompleteness theorems computers necessarily have to process incomplete data, thus necessarily giving rise to consideration of the semantic concept of *partial information.* Mathematics before the 1960s had no ready made answers on how to model information which in a well defined sense had bits missing. Computer scientists wishing to understand the semantics of programming languages had to find ways to *denote* information which could, quite legitimately & usefully, be incomplete. An analogy can be drawn from mathematics where there are many applications requiring us to have a real number i such that $i^2 = -1$. As clearly no such real number exists we creatively extend the real numbers to include appropriately termed *imaginary* numbers, one of which is to be $i = \sqrt{-1}$. The mathematical solution for denoting *partiality* accords a pivotal role to *nothing*, that information totally devoid of all content. In what sense then can nothing be said to exist? Posed as an oxymoron, what is the sound of silence? Fortunately we have available to us the inspired mathematics developed by, largely one person, Dana Scott to resolve the potential logical paradox of self reference embodied in any non trivial deterministic programming language. Roughly speaking a computer program is *deterministic* if each time it is executed the same output necessarily results. What are the implicit philosophical premises upon the nature of information assumed in Scott's ground breaking technical work? And what do these premises tell us about why his work has great difficulty in generalising to *non deterministic* programs, those where one execution may legitimately yield different results from another. In this paper we take a non deterministic program to be one where at one or more points in the program's execution the computer may choose any one from a finite set of commands specified in the program to execute next. A non deterministic program thus specifies a possibly infinite number of possible execution sequences. The work in this paper is an attempt to identify the key premises upon the fundamental nature of information underpinning Scott's work, and to demonstrate how computer science has successfully generalised the ontology of information inherited from mathematics. This paper is intended to promote research into ontological backdrops for existing practical work on the epistemology of reasoning, knowledge representation, and knowledge acquisition using computers. A clearer understanding of such ontology could make existing highly technical work on the semantics of programming languages more accessible to programmers. For philosophy, this work is intended to promote debate upon how we might ultimately develop a *theory of everything* to reconcile the essence of mathematics with that of computer science.

Premises for partiality

We now attempt to identify the implicit premises in Scott's approach to modelling partial information in the semantics of programming languages.

Premise 1 : programming language semantics is certain knowledge.

Assigning a semantics to a program is the problem of determining and expressing all of its *certain* properties, these being properties which are provable in an appropriate logic of programs. By implication, there is no room here for approximation, lack of clarity, or ignorance of program properties. A program is *understood to be exactly what it is, neither in part of what it is, nor what it might be.*

Premise 2 : a program's semantics is mathematically denotable.

This premise asserts that a mathematical model can be constructed in which each program can, for its meaning, be assigned a single value in that model.

Premise 3 : the semantics of a program is its observable behaviour.

The semantics of a deterministic program is the totality of effects which can be observed and recorded (by a human or another program) in the one, and only possible, execution sequence. At the risk of using an anthropomorphism, we may say that a deterministic program *is* what it *does*[4]. In contrast a non deterministic program may, when executed, produce one of many possible execution sequences. Thus, what may we say a non deterministic program *does*? Attempts have been made to generalise Scott's approach to non deterministic programs, so providing, in accordance with premise 2, a denotation for non determinism. By premise 1 we require certain knowledge of all properties of each non deterministic program. Thus, the meaning of a non deterministic program has to entail all the certain properties of each and every possible execution sequence. Thus, non determinism, it is reasoned by some, can be captured as a set of determinisms, a *determinism* being the meaning assigned to a deterministic program. And so, by premise 2, we need a mathematical model in which a value is a set of determinisms. And so, if S is the set of all possible determinisms for deterministic programs, then all we should need is a so-called *power domain* 2^S of subsets of S, each such subset to be the meaning assigned to a non deterministic program. The power domain approach presumes non determinism to be a set of determinisms. Non determinism is defined in terms of the primitive notion of *determinism*, and as such is a kind of, what we may say, *multiple determinism.*

Process calculi such as Milner's *Calculus of Communicating Systems* (*CCS*) [Mi89] have presumed non determinism to be a primitive notion, rather than a derivative of determinism. The process $P + Q$ (pronounced *P or Q*) is the process which can either behave like process P or behave like process Q. '+' is introduced as a primitive notion of *choice.* While the program may specify what choices there are, such as a coin has the sides *head* and *tail*, it is the computer executing the program, as in the tossing of a coin, that will make the choice. *CCS* reverses the approach of power domains, by asserting choice to be a core primitive notion, and that determinism is a process having just one choice. In other words, determinism is a choice of one. And so, in accordance with premise 2, what is the denotation of choice? An equivalence relation[5], termed

[4] The use of the terms *does* and *can do* in this paper are common parlance in computer science. They have no anthropomorphic connotations, they merely refer to the underlying capabilities of the computer system used to execute programs.

[5] A binary relation $\equiv$ over a set A is an *equivalence* if it has the following properties for all a, b, c in A. $a \equiv a$. If $a \equiv b$ then $b \equiv a$. If $a \equiv b$ and $b \equiv c$ then $a \equiv c$. The *equivalence class* for a is the set of all those members of A equivalent to a.

a *bisimulation* (pronounced *bi simulation*), on processes whose semantics are to be regarded as logically equivalent is introduced. The denotation of a process is its equivalence class.

Premise 4 : the semantics of a program is precisely that information which can be discovered by means of Turing computation.

This premise asserts that we can know whatever the computer can reveal to us by way of (Turing) computation, but no more. It is thus, by presumption, not permitted to enhance our understanding of the program's meaning by use of clever reasoning which would go beyond the Church-Turing hypothesis universally accepted in computer science as the definition of what is computable. The semantics of a program is thus necessarily taken to be precisely the computation(s) it describes. For a deterministic program we observe what it *does*, as this word was used earlier, in its one and only possible computation. For a non deterministic program, such as a process in *CCS*, we need to observe both what choices are available and what each such choice (when chosen) *does*. The required notion of *observation* is thus twofold, what a process *can do* and what it *does*. As a deterministic program only *can do* what it *does*, the notion of *can do* is safely dropped from consideration in its semantics.

Premise 5 : nothing is observable.

The problem of how to assign a denotation for a non terminating loop in a deterministic program written in a typical Pascal-like language loosely corresponds to the problem of assigning a semantics for recursive function theory, which in turn is similar to the problem of how to demonstrate the logical consistency of self applications such as $x(x)$ in the lambda calculus. The initial and instrumental technical concept in Scott's work is to utilise the notion of $\perp$ (pronounced *bottom*). $\perp$ can be understood for Pascal-like languages as a non terminating program which, while remaining alive, never progresses in any observable sense of producing further output. For example, the following Pascal code will run forever, but never produce any output.

```
while true do begin end ;
writeln("hello") ;
```

In other words, silence is the 'output' of the above code that never outputs anything. As with a person in a coma with irreparable brain damage, this program will remain *alive* indefinitely when executed, yet will never usefully progress in the sense that the `writeln` command is never executed, and so cannot produce the observable output `hello`. $\perp$ is introduced as a denotation for that which is undefinable. We can know all there is to know about $\perp$ up to the extent to which $\perp$ is defined. In summary, Scott cleverly arranges in his work that any entity will exist only up to the extent to which we can know, in accordance with Premise 4, its properties by means of computation. Consequently the meaning of something defined in terms of itself is that information which can be computed from that definition, and no more. In summary, $\bot$ creatively introduces the intriguing oxymoron, *the sound of silence* as the starting point for computing a meaning for self reference.

So, what is *nothing* in a non deterministic language such as a process calculus? Hoare's *Co-operating Sequential Processes* (*CSP*)[Ro98] has a denotational semantics, a refinement of the power domain approach. If, as power domains do, we can take sets of determinisms, why not take sets of partial determinisms as such partiality is understood in Scott's work. *CSP* does just this, it uses, what we may term, *multiple partiality* to construct a Scott-like semantics for non determinism. Just as power domains define non determinism in terms of determinism, so *CSP* understands non determinism to be defined in terms of partial determinism. Power domains and *CSP* thus share the notion of non determinism as being derived from a notion of determinism, the former from totally defined determinisms, the latter from partial determinisms. This is not a reconciliation of determinism with (say) the primitive notion of *choice* as used in *CCS*, but a wishfully simplistic reduction of non determinism. Such a reconciliation is, we argue here, necessary if, in accordance with premise 5, the nothingness which may be either or both of *does* $\bot$ and the (primitive) choice of *can do* $\bot$ is to be truly observable.

A technical fix for an ontological problem?

Bisimulation, undoubtedly ingenious[6], is ultimately a mathematical device for avoiding an ontological problem. We have denotational approaches which can model non determinism in terms of total (i.e. power domains) or partial (i.e. *CSP*) determinism, and a non denotational approach (i.e. *CCS*) having choice as a primitive.

Our suggested premises trace the initial steps in Scott's mathematical constructions. More such premises would be needed for a complete treatment of his work, but the subject of this paper is to discuss philosophical issues surrounding concepts such as $\bot$, *partiality*, and *can do* involving less than certain knowledge of programs. In premises 1 through 5 we have established the key concept of $\bot$. $\bot$ is the starting point for a comprehensive mathematical theory to provide a denotation for each program in a deterministic programming language[AJ94]. $\bot$ is at first sight contradictory, it is the value defined for that which is totally undefined, an ingenious mathematical device to reason about that which is, in a well defined sense, unknown. There is no problem of self reference as the theory is carefully configured to ensure that we know partiality up to, but no more, the extent to which it is known with certainty by means of computation. Mathematically this works just fine, however, it will not generalise to include *choice* as a primitive notion. In contrast the process calculus *CCS* can elegantly handle

[6] The importance of bisimulation, created by Park & Milner for *CCS*, was clearly demonstrated when subsequently the mathematician Peter Aczel defined a theory of non-well-founded sets[Ac88], allowing for a set to have an infinite nesting of subsets. For example, the recursive definition $A = \{A\}$ defines a non-well-founded set $\{\{\{\ldots\}\}\}$, but this would not be a legitimate set in a well-founded set theory such as Zermelo-Frankel.

non determinism using a bisimulation relation, but, for this to work, excludes partial objects such as $\perp$. Computer science semanticists, studying either determinism or non determinism are agreed upon the validity of premises 1, 2, & 3, thus leading to a stark choice of either a denotation of programs having an overly simplistic reduction of non determinism to determinism, or a relation upon programs without the deterministic $\perp$. Is there no model of reconciliation? Is there no model in classical logic having *referents* for certain knowledge of programs such as deterministic behaviour, yet can accommodate the inherent lack of certainty which is non determinism? There is a well known problem in philosophy which strongly suggests that such searching is in vain. In addition, research into this problem indicates an alternative way forward, one which leads us to challenge the validity of the *certain knowledge* of premise 1.

Vagueness

Williamson[Wi96] describes the *Sorites paradox* as being one of seven puzzles proposed by the logician Eubulides of Miletus. No one disputes that that one grain of sand does not constitute a *heap*, and likewise that a trillion does make a heap. So, how many grains are needed to constitute a heap? Williamson traces the development of this perplexing problem from ancient Greece to fuzzy logic[Kos94], concluding that *"... none of the alternative approaches has given a satisfying account of vagueness without falling back upon classical logic"*. In one way or another there is a presumption that we can obtain with certainty a knowledgeable solution to the paradox, and so the inevitable recourse to the only certain language which is logic. Programming language semantics makes the same assumption, that each program can be all known, hence our premise 1. In contrast the serious programmer, such as any of our computer science undergraduates here at Warwick, 'know' that all the 'horrible' mathematics taught in our semantics course is even more complicated than the task of programming itself! They 'know' that one designs good software to be *reliable*, that is, to have a very high chance of producing desired results. They 'know' that producing a *totally correct* program in accordance with a semantics given as a specification (in an appropriate system of mathematics or logic) of what it can or should do is in practice, if not in theory, usually beyond their reach. The notion of *reliability* to a programmer is as a *heap* is to a philosopher. A program that always crashes when executed is clearly not reliable, while one that has run numerous times without a problem clearly is reliable. So, what is *reliability*? Computer science semanticists are sadly disdaining of the notion of reliability, only willing to discuss the certainty of total correctness. Philosophers have to their credit struggled with the notion of *heap*, while computer scientists have, through premise 1, had to reject reliability, a common sense notion used by all accomplished programmers. Reliability is a notion of vagueness that programming language semantics has yet to embrace, as the Sorites paradox has been embraced in philosophy.

The serious programmer does not expect to 'know' everything about their program, accepting that many of its properties cannot, for reasons they care not of, be known. The argument in this paper is that the insistence upon certain knowledge of premise 1 forces us to seek certain models of uncertain situations such as non determinism. In contrast, fuzzy logic asserts that an imprecise truth value can be modelled by a precise real number between 0 and 1. Following Scott's example in the characterisation of $\perp$, we argue for the following position, contrary to those of both premise 1 and fuzzy logic.

Premise of necessary uncertainty : the partial can at best be known up to the extent to which it is partial.

This premise implies the existence of *ignorance* in programming language semantics, the inclusion of *necessary uncertainty.* Not only may our knowledge of programs be partial, as in the case of $\perp$, but in addition our ability *to know* may be partial as in the case of the actual choices made during the execution of a non deterministic program. A computer program to simulate the tossing of a coin can specify in its code the two possibilities of *head* and *tail,* what the program *can do.* What the program *does* when executed cannot be known from the program's semantics. Yet, paradoxically, current denotational models of non determinism, such as those for *CSP,* define *can do* in terms of *does,* when no one can know what *does* happen until it has happened. The result is that current work on semantics is really *can-do-semantics,* the only knowledge that can possibly conform to premise 1, and as such is of little use to the serious programmer whose common sense mind visualises what the program *does.* The prevailing idealistic culture of mathematical certainty in programming language semantics severely inhibits communication with programmers whose need is for a usable model of reliability. Program reliability is not a notion in semantics that classical mathematics can model as it is not a matter of certainty; mathematics needs to work with vagueness.

Williamson's thesis is that vagueness is, *"... an epistemic phenomenon, a kind of ignorance: there really is a specific grain of sand whose removal turns the heap into a non-heap, but that we cannot know which one it is"* He makes the thoroughly realist point that, *"... even the truth about the boundaries of our concepts can be beyond our capacity to know it"*

Conclusions and further work

This paper has studied a philosophical issue in computer science unrelated to the traditional problem of whether or not a computer can *think.* The usual differing technical approaches to determinism and non determinism in the semantics of programming languages have been studied as an example of a misrepresented ontological issue, and subsequently compared to the problem of vagueness in philosophy. Following Williamson's treatment of the Sorites paradox, we have argued that a reconciliation of determinism and choice has to relax the traditional presumption for the certainty of knowledge of program properties.

Scott's notion of partiality is traditionally modelled in a (point set) topology by weakening the Hausdorff (i.e. T_2) notion of separability, usually assumed in mainstream mathematics, to T_0 separability. Such T_0 topologies have subsequently been modelled using a form of generalised metric topology[Ma95], leading naturally to the study of bi-topology[Kop04]. A *bi-topology* is a pair of related topologies over the same universe of points, thus suggesting the inadequacy of a single topology to model necessary uncertainty encountered in programming language semantics.

The possibility of a necessary separation of a topological space from our knowledge of it is unknown in the reductionism of classical mathematics, a school of thought to which denotational semantics has always been strongly affiliated. This has left semantics isolated from the potential benefits of 'inherent uncertainty' that have been embraced by chaos theory and quantum computing. Computer science, quantum physics, and philosophy, each in their own distinct way, suggest that a separation of *object* from *knowledge* can usefully, perhaps necessarily, be drawn between the object of study and our capacity to know it. A bi-topology could serve to model one topology defining the object of our study, and a second to tell us what we may be permitted to know in reasoning about the first.

The authors' researches so far[Ma95,Ma02,Kop04], observing as they do the doctrine of premise 1, have nonetheless made useful progress in advancing our technical understanding of *partiality*. But, further progress appears to require a definitive separation of *object* from *knowledge*. Ours, and any other related work, needs to accommodate the possibility of necessary uncertainty. Both *partial metric topology*[Ma95] and Williamson's *logic of clarity*[Wi96, Appendix] achieve such accommodation by quantifying the extent of partialness. At a more fundamental level philosophers need to engage with computer science and mathematics on re-interpreting the classical puzzle of the Sorites paradox as an epistemological problem of necessary uncertainty in computing.

References

[AJ94] S. Abramsky & A. Jung, *Domain Theory*, in *Handbook of Logic in Computer Science*, Vol. 3, Clarendon Press, pp.1-168, 1994.

[Ac88] P. Aczel, *Non-well-found Sets*, CSLI/SRI International, Menlo Park, CA94025, 1988.

[Kos94] B. Kosko, *Fuzzy thinking: the new science of fuzzy logic*, Flamingo, London 1994.

[Kop04] K. Kopperman, S. Matthews, & H. Pajoohesh, *Partial metrizability in value quantales*, Applied General Topology, to appear 2004.

[Ma02] S.G. Matthews, *Pixel geometry*, Electronic Notes in Theoretical Computer Science, vol.40, `www1.elsevier.com/gej-ng/31/29/23/show` , 2002.

[Ma95] S.G. Matthews, *Partial metric topology*, in *Papers on general topology and applications*, Proc. Eighth Summer Conference at Queen's College, Annals of the New York Academy of Sciences vol.728, pp.183-197, 1995.

[Mi89] R. Milner, *Communication and concurrency*, Prentice Hall, London 1989.

[Ro98] A.W. Roscoe, *The theory and practice of concurrency*, Prentice Hall Series in Computer Science 1998.

[Wi96] T. Williamson, *Vagueness*, *The Problems of Philosophy* Series, Routledge 1996.

Reconcile art and culture on the Web : Lessen the importance of *instantiation* so creation can better *fiction*

Francis Rousseaux[1], Alain Bonardi[2]

[1] Laboratoire d'Etudes et de Recherches en Informatique
University of Reims Champagne Ardennes (URCA)
IST SemanticHIFI European project team
Institut de Recherche et de Coordination Acoustique/Musique (IRCAM)
1, place Igor-Stravinsky
75004 Paris France
francis.rousseaux@ircam.fr
[2] Artificial Intelligence and Mobile Robotics Research Team
University of Paris 8
Maison des Sciences de l'Homme Paris-Nord
4, rue de la Croix-Faron
93210 La Plaine Saint-Denis France
alain.bonardi@wanadoo.fr

Abstract. In this paper, we show how the practice of buying CDs surreptitiously conditions our musical activities, and how the slowly-evolving practice of classifying structured by buying and the notion of genre, will disappear, as develops an *ad-hoc* organization of auditory samples, centered on prototypes and similarity. Thus appears similarity-based calculus, acting on considerable masses of samples, which continuously change the balance of the man-machine dialogue, in an appeal started again and again to compare information. We also show how some artistic creations proceed in the same way. Art and culture are linked in processes that take advantage of their massive digitalization.

1 Preliminary Notes

Instantiation is often used by computer scientists, which comes from the word *instance,* which means *example, case;* instantiation somehow generalizes the operation used by mathematicians by which a numerical value is assigned to a variable. To speak about reality, computer scientists instantiate abstract categories, thus decreeing that this or that entity is a specific instantiation of a category, which itself is linked to other categories by general hierarchies and/or formal properties. The whole device [10] is sometimes called an *ontology* —ontologies are assumed to describe sections of mundane knowledge widely used in artificial intelligence, sometimes an *object-oriented design.*

The neologism *to fiction* is used here to remind us that an interactive computer device based on symbol manipulation *works* only if the two following conditions are met: on the one hand, that programs starting an execution have a correct syntax, and on the other hand, that the user agrees to play the interaction game with the device, thus recognizing it as an operational fiction.

2 Navigating in oceans of music: a fascinating indexing problem

Using a classifying mode always implies more assumptions than it seems. And this is why the issue of classifying is an old philosophical issue [1], [11].

2.1 Today's craze for on-line music distribution

If online music distribution has caused a lot of ink to flow lately [13], it is because three circumstances allow social, cultural and business demands to persist and crystallize, and even to be so important as to become strategic industrial necessities:

- people who buy CDs do not understand that the Web is still deaf to their appetite for music "à la carte" [12];
- music designers fight to escape the normative constraints implied by CDs;
- music producers and music companies are organized in a small number of multinational companies in competition (the famous *Majors*), and they fear that a major innovation will disrupt their balance.

We will limit our study to the investigation of the *indexing of musical objects*, that is to say the study of the organization that would enable accessing and listening to music in a differentiated way, through the Web as a matter of course.

2.2 CDs surreptitiously condition our musical activities

To begin, let's consider the world of CDs: they are concrete materials which contain sequences of musical pieces, and buying these concrete objects on the market is a necessary step to access their content.

To be able to sell CDs, it is necessary to arrange them on displays, and to label the shelves and the trays. It is not obvious at all, but the decisions taken here—very often implicitly—on the way CDs are organized, will have a considerable influence on the way our musical activities are established and organized, and especially on music listening, which is supposed to be so ineffable!

In music stores, each tray is labeled to allow a meaningful organization of products, by making a compromise between the physical constraints of the store, the clients buying habits, and the label readability.

But the compromises made by department managers of record dealers rival each other in cleverness and originality—this is how some items are sometimes present in *several different trays* at once. Faced with such compromises, the promoters of the science of classification would often tear out their hair. But business efficiency prevails in this case over scientific rigor, and systematic coherence matters less than productivity [8]!

It is indeed the *buying* activity which conditions the indexing mode of CDs on the shelves of CD stores. When a buyer recommends a CD to a friend who will potentially buy it, he will explicitly use the label categories, which semiotic system he has learnt without knowing it, by walking around wholesale dealers...And the labeling device soon becomes part of a cultural heritage; it won't be long before it is used to describe the whole range of musical activities, including music listening itself, affected even in its most meditative modality [16].

2.2 From CDs to digital sound files

Some think that forcing descriptive labels into digitalized online distributed *sound files*, is a good way to elaborate online music distribution services. These labels then become what is called *meta-information.*

This solution is not devoid of interest for the comfort and the cognitive ergonomics of the users of online services, because it allows them to question machines in the way they questioned labels not long ago. Walking around spaces cluttered with shelves is here simulated by using pop up menus and lists on a screen. New problems will certainly arise when downloaded files will be organized, essentially from file names [9].

More generally, those who promote these solutions are not unaware of their built-in drawbacks: they know that the price to pay to maintain and update this meta-information is high, that chances are weak that we will able to extract them automatically from musical data, and that the labeling system itself must first be rationalized to become compatible with a computerized processing.

But they often fail to understand the most reactionary aspect of their viewpoint. If we follow Gilbert Simondon and his theoretical proposal on the existence mode of technical objects [14], when mechanisms of online music distribution will come into being, they will inexorably be far from these inevitably temporary solutions ...

Drawing our inspiration from Simondon, we can hypothesize that with the loss of interest for CDs and the end of the requirement to buy the medium *before* listening to music, we will witness the corollary fall of an operational fiction, that was simplistic, but effective nonetheless. The object of the fantasy will not be to access the CD anymore, but the piece of music, or the sample, in a way that we must clarify. Because other forms of musical activities will emerge to make possible the organization of musical objects, which will be more often dedicated and signed. Thus the normative activity of CD buying will give way to situated actions, that will be part of more and more differentiated and singular projects.

2.3 The object/activity/description triad

This is what we meant when we wrote about "music-ripping" [3], to mean that listening means signed listening/composing/producing, based on *samples*.

What is the array of listening situations? Will it be necessary to try to define it by referring to professional or amateur practices—the way a composer listens when he is trying to compose, the way anyone listens to music in the shower in the morning, etc. ? No, because these stereotypes pertain too much to the ancient era and say nothing about the world of listening as signed [5], that is to say about listening when it causes appropriations and mediations (digital ones are what matter to us).

Therefore, other typologies of signed listening situations will emerge shortly, in so far as they will be facilitated by the emerging technical systems of indexation and navigation [17] which will stimulate us to imagine them. The slowly-evolving practice of classifying *a priori*, structured by buying and the notion of *genre*, will disappear, as develops an appropriate organization of auditory samples, centered on *prototypes* and *similarity*, situated in the singularity of everyday practice.

3. Using calculation though staying the master of the game: lessen the importance of instantiation

How is it possible to inspect a singularity in normative category systems and calculation procedures? These devices are not arbitrary—they have their own life, their individual and concrete nature, but the *relationship* between the singular and the particular linked to generalities could very well look arbitrary. In computer science, this is what *instantiation* is, though it has never been described, and its effective reduction probably implies, as an *infinite remedial task,* consequences on the systems of categories themselves and their individuation.

To be able to make sense of the effectiveness of calculation, singularities of the living world must have been *instantiated* with symbolic conceptual particulars, which will be enlisted in conceptual systems (semantic networks or other ontologies) meant to represent knowledge relationships and to make their approximation possible by calculable models.

If it is reasonable to consider these conceptual devices as technical abstract entities, which as such will possibly materialize (in Simondon's sense), and become more refined —also by integrating more and more sophisticated meta-models, it is probably unreasonable not to see that their materialization is determined by the gap between 1° what they pretend to represent in order to make calculation possible (finalized vision) and 2° what they tend to recapture from the roughness of the instantiation operation (original vision).

3.1. Instantiation as a calculation *incitement*

Instantiation is indeed the part of computer science that has not been reflected upon. Better: instantiation is the fundamental usurpation of computer science, its radical cheating, its dead angle. This has serious consequences on thinking and research in computer science.

First, the spontaneous assimilation of singularities to particulars linked to devices of conceptual meshing leaves the whole task of representing and simulating reality to these devices. This is how research in computer science wears itself out setting up devices, forgetting from the start that their specifications are prescribed underground by the attempt to compensate for the obscurity of the operations of instantiation. This probably explains the infinite development of research on the representation of knowledge and ontologies, which aims at materializing (by category differentiation) devices that are supposed to be able to give light to a black hole.

Second, the consequence of the practice of computer scientists is often to lessen the importance of the operations of instantiation, as if to veil its un-reflected nature. Even if it is rare for a computer scientist to claim this inspiration explicitly—Frédéric Drouillon [6] however places it at the heart of his creative work, by assuming that is possible to shy away from instantiation by stealing already instantiated systems to enroll them in more complex systems, it often operates as background to research [15].

Among the implicit means that enable designers of computing systems to lessen the importance of instantiations, there is one that relies on a heuristic meaning of calculation and on a vision of the request-calculus interactivity which is on the edge of contradicting Turing's hypotheses on the investment of the machines' minds by dialog, which is at the source of their reputation of intelligence. In the context of searching for digital content using similarity, this means is usually used.

Because when a calculation is used in order to organize a great number of digital entities according to their similarity to a specific example, the symbolic interpretation of the calculation and of the symbols that are used for it must be given up, and the only possible outcome is its heuristic efficiency in selecting particulars, proposed as so many candidates which undergo a singular election to be accepted by a user. This is how the fundamentally heuristic nature of the resulting similarity-based calculus is expressed—for any singular demand, there is always a corresponding instance among the many particular candidates.

In this case, the multi-criteria instantiation is done conforming to a distance, which allows a strong semantic interpretation even less as it relies on different means of statistical comparison, which retain the coherent and reproducible behavior of a user, or of a community of users. The important gesture is still the choice of the user when he recruits a singularity among a pre-selection of particulars. And it is only in this way that calculation is interpreted, in an "interested" way.

It is necessary to add that the user does not have to make a definite choice, but that on the contrary, he is encouraged to perform again his gesture of heuristic enquiry by similarity, starting with a new singularity and if necessary with new specifications criteria of the desired distance. Therefore, in repeating the gesture of aided selection, an evanescence of instantiation continues, sliding from selection to selection. And the dialog with the calculation is not part of Turing's intuitions, somewhat similarly to how the balance of a "balanced" chemical reaction can always be displaced by withdrawing progressively the material resulting from one of the two ways of the reaction. Balance is meta-stable. The other side of calculation is of no interest to the user who does not try to build its model, but only to profit from it dynamically.

3.2 Heuristic and meta-stable similarity: a growing mobilization of calculation

Let's come back to the example of online music distribution. The activity of buying CDs used to condition the indexation system. The activities around searches for on-line data will focus on ad hoc nuclei and idiosyncratic goals, which will closely unite reception and action,, and thus determine their own objects, which we call samples [3].

The *a priori* indexation is gradually replaced by a dynamic indexation, and its central paradigm is usually similarity. This implies a need for algorithms to calculate descriptors on the basis of digital contents themselves, and an attempt to match these (self-extracting) descriptors to appropriate categories in the context of committed *ad hoc* activities. Indexation devices can then be composed of meta-information, but also of labels calculated from musical contents. The MPEG7 and MPEG21 norms thus propose to put on the market descriptors created from automatic extraction procedures and names that identify descriptive qualities.

4. Drama Interlude: two ways of using instantiation in the intermedia play *La traversée de la nuit - Through the night*

The theater play[1] *La traversée de la nuit* by Geneviève de Gaulle [7] is the story of the author, imprisoned in the dungeon of the concentration camp in Ravensbrück.

4.1 Instantiating with neural networks the expression of a comedian on stage

The work on memory is here linked to a neural network, which is used to recognize the emotional states from the voice of the comedian who says the whole text. Neural networks are a clever and fashionable way to conceal instantiation.

For a few months, the comedian Valérie Le Louédec declaimed her text in front of the computer, in preset moods: joy, sadness, resignation, anger, etc. After choosing a mood, the comedian tries to say the text in that state, no matter what the meaning of the text implies. For every pronounced sentence, a vector of twelve components is extracted from her voice: four for the vowels, four for the consonants, four for the prosody. Then these vectors need to be linked to the intended emotional state, and this is the learning phase for the neural network.

Thus, alternatively to an explicit model of an emotional state, a neural network is meant to set similarity conditions statistically. The process of establishing similarities, by constituting equivalence categories, is based on the massive repetition of the

[1] Play performed on November 21st, 22nd, and 23rd at the Center for the Arts at Enghien-les-Bains (95). Director: Christine Zeppenfeld; comedians : Valérie Le Louédec and Magali Bruneau; multimedia design : Alain Bonardi and Nathalie Dazin; music : Stéphane Grémaud; lights : Thierry Fratissier. Website : http://www.latraverseedelanuit.net

experience of the phenomenon. When the comedian is rehearsing or acting, the neural network "makes every effort" to recognize immediately the states of the comedian's voice. The quest for similarity continues.

In the two cases, learning and recognition, there is not any real dialog man-machine in the traditional sense of question and answer, but a frantic solicitation of one by the other. While learning, the neural network constantly appeals to the comedian; during recognition, the comedian is the one who appeals to the network relentlessly, to adapt her acting to what is stable and what changes. In both cases, one does not listen to the other, in the classical meaning of a dialog, but one works by continuously taking in information given by the other one.

4.2 Instantiating a collaborative graphic generator

The main goal of the device in this play is to make the comedian's voice control the character's "mental images," which are projected on the screen at the back of the stage, on a vast cyclorama (30 by 16 feet). The demands of artistic expression thus lead the designers to lessen the importance of instantiation, to give up the classical way of specifying a graphical problem.

Therefore autonomous agents were used to build the picture at the back of the stage in a collaborative way. These agents should be pictured as billposters who would work together to create a poster from fragments of images along with the sponsor's goals. The prupose is to create an artistic content by solving iteratively an optimization problem.

Each of these billposters is modeled according to a few variables, inspired by psychology, which correspond to different states of the voice (recognized by the neural network) which increase or decrease its "mood".

The results produced by this generative device that is given goals are amazing in terms of the distortion and movement of the image. Across performances, the movements on the screen are completely different, but they are always somehow "harmonious". This technical invention has also changed the relationship between the computer and the members of the project.

Some drama categories are thus displaced or questioned by this way of accessing artistic expression according to similarity and repetition. For a long time, the text of the play was long given as the only input towards expression and emotion [2].

Doing away with the text as the only medium, and especially by using the medium of digitalized voice, the processes that we have described, based on the establishment of similarities, pave the way for the creation of new drama emotions, and maybe new ways of theater creation and distribution with digital devices, provoking new uses for it. As with the distribution of musical contents by the Web, the auditory signal, its descriptors, the organization of "drama" contents with databases and MPEG norms that could be imagined, would reorganize the connection between theater and the audience[2].

---

[2] We have partially studied this kind of reorganization in the case of digital opera, see [4].

Conclusion

We have shown how the distribution of digitalized music does away with the dominant paradigm based on buying CDs for a dynamic indexation centered on the notion of similarity. Thus appears similarity-based calculus, acting on considerable masses of samples, which continuously change the balance of the man-machine dialogue, in an appeal started again and again to compare information. We have also shown how some artistic creations, heavily based on computers, proceed in the same way.

These prospects open many research avenues, building bridges especially between the search for content and artistic creativity. Great disruptions in the domains of culture and art, now intertwined by their massive digitalization, can be foreseen.

References

[1] Aristote, *Métaphysique,* Garnier Flammarion, 1993

[2] Beaurivage, Th., *Exploiter des données morphosyntaxiques pour l'étude statistique des genres – application au roman policier*, "Traitement automatique des langues" Journal, vol. 42, 2, text available at www.atala.org/tal/, 2001

[3] Bonardi, A., Rousseaux, F., *Music-ripping : des pratiques qui provoquent la musicologie,* Musicae Scientiae, 2004

[4] Bonardi, A., Rousseaux, F. *Composing an Interactive Virtual Opera : The Virtualis Project*, Leonardo Journal, Journal of the International Society for the Arts, Sciences and Technology, vol. 35, 3, p. 315-318, 2002

[5] Donin, N., *Towards Organised Listening: Some Aspects of the Signed Listening Project at Ircam,* Organised Sound, Cambridge University Press, forhtcoming in 2004

[6] Drouillon, F., *Le calcul de la beauté : le principe de Pinocchio,* proceedings of the STP seminar, MSH Paris, 2003

[7] de Gaulle, G., *La traversée de la nuit,* Éditions du Seuil, 1998

[8] Pachet, F., *A Taxonomy of Musical Genres,* RIAO, Paris, avril 2000

[9] Pachet, F., *Nom de fichiers : Le nom,* proceedings of the STP seminar, MSH Paris, 2003

[10] Perrot, J.-F., *Des objets aux connaissances,* Journée Méthodes objets et Intelligence Artificielle : Frontières, Ponts et Synergies, Paris RIA, juin 1994

[11] Rastier, F., *Sémantique pour l'analyse,* Masson, Paris, 1994

[12] Rousseaux, F., *Pourquoi le Web resterait-il sourd à nos envies de musique et de son ?,* MEDIANET, Monastir, 2002

[13] Rousseaux, F., *La notion d'œuvre à l'heure de la musique en ligne : comment concevoir la nécessaire médiation des affinités électives ?,* Klincksieck Editions, 2002

[14] Simondon, G., *Du mode d'existence des objets techniques,* Aubier, 1989

[15] Seridi, H., Akdag, H., *Approximate Reasoning for Processing Uncertainty,* Journal of Advanced Computational Intelligence, vol. 5, 2, Fuji Technology Press, 2001

[16] Szendy, P., *Ecoute, une histoire de nos oreilles,* Éditions de Minuit, 2001

[17] Vinet, H., Herrera, P., Pachet, F., *The Cuidado Project,* proceedings of the ISMIR Conference, Paris, 2002

Self-referencing languages revisited

Gábor Rédey[1] and Attila Neumann[2]

[1] Hungarian Atomic Energy Authority
H-1036 Budapest, Fényes A. u. 4., Hungary
[2] Technische Universität Wien
Institut für Computergraphik und Algorithmen
Favoritenstraße 9-11 / E186, A-1040 Wien, Austria

Abstract. Paradoxes, particularly Tarski's liar paradox, represent an ongoing challenge that have long attracted special interest. There have been numerous attempts to give either a formal or a more realistic resolution to this area based on natural logical intuition or common sense. The present semantic analysis of the problem components concludes that the traditional language of logic fails to detect Tarski's paradox, since the formalised version of the liar sentence does not represent a correct definition. Neither the formal language, nor the logical system is deficient in this respect. Only natural language statements cannot be interpreted adequately by traditional language of logic.

1 Introduction

Paradoxes play a remarkable role in philosophy and logic. Several have been resolved, eliminated or brought to rest by appropriate theories (e.g.: Cretan paradox: Epimenides the Cretan says "All Cretans are liars", or the key statement of Nihilism: "there is no truth"), while others attract perpetual and changeless interest. Time and again new researches they challenge, and consistently elude resolution.

One of the most impressive paradoxes is analysed in this paper, specifically Tarski's liar paradox associated with the definition of logical truth. Much effort has been made since Tarski's theory of meta languages to obtain a semantically more acceptable explanation. For a thorough, complete and even historical analysis of this matter the reader can be referred to Feferman [2], although, there are many other papers continuously re-examining this point [1, 4, 5, 3].

Self-referencing languages surely imply a fundamental dilemma within philosophical logic. However, this property also demands particular interest from the Artificial Intelligence community. Clearly, a potential attribute of representation languages is beneficial in reasoning systems of any kind, no matter whether it is for a human or robot. Natural languages are essentially rich in introspective statements concerning feelings, remarks, opinions, knowledge or other content-related features. Assertions of this kind apply a large variety of linguistic devices such as indirect quotations or modalities which are also occasionally loaded by self-reference. Therefore, an adequate representation of self-reference is crucial not only theoretically or philosophically, but also from a practical, functional point of view.

The problem occurs not at the level of natural languages, but that of the formalisation process. Formalisation of some natural language sentences can be inadequate even if it is possible. The latter alternative can obviously be ignored, since it would impugn the possibility of natural human understanding, without which there is no meaningful communication, formal or otherwise. The justification and resolution of the former option is the objective of this paper.

The foremost exact phrasing and modelling of the informal liar sentences was originated by Tarski and subsequently by numerous alternatives by Kripke, Gilmore, Feferman, Perlis, Kerber and others. It will be shown that the initial formal rephrasing of the liar sentence does not satisfy the natural requirement of being a valid definition.

The paper presents an attempt to answer this problem, which is based on a different logical language, determined by one of the authors, which has been proven to have a better expressive power.

2 Traditional formalising self-reference

The problem itself cannot even be identified without any formal representation. The first realization relating a Cretan type of sentence is accomplished by Tarski.

2.1 Tarski's paradox

Roots of Tarski's paradox are summarised here from [2]. Let $x, \ldots$ range over the statements of the language of a logical system, which are assumed to be closed under the usual propositional operators denoted by $\sim$, &, $\vee$, $\supset$, $\equiv$. Each statement x of the language has a name, i.e. there is an associated closed term $\ulcorner x \urcorner$ of the language. Then the following axiom is accepted for a predicate $t(\ulcorner x \urcorner)$, which is interpreted as expressing that x is true:

$$t(\ulcorner x \urcorner) \equiv x \tag{1}$$

for each statement x of the language.

For the derivation of a contradiction in this system the liar sentence is taken:

$$\text{(2) \textit{is not true.}} \tag{2}$$

formalised mostly as

$$l :\equiv \sim t(\ulcorner l \urcorner) \tag{3}$$

By definition (1) and transitivity of biconditional we obtain $t(\ulcorner l \urcorner) \equiv \sim t(\ulcorner l \urcorner)$, immediately implying inconsistency of the system.

2.2 The solution routes

There has been considerable work on theories eliminating these paradoxes. Numerous, seemingly different escape routes have been determined the common idea of which is to discard problematic sentences:

1. either by altering the syntax so that undesirable statements could be excluded from the language,
2. or by revising critical axioms semantically so that antinomic formulae could be evaluated exceptionally (e.g. as meaningless ones).

Because of limited space, individual instances of the above categories are not discussed here. The former is followed by Tarski and Kerber [3], while the latter is preferred by Feferman [2] and Perlis [4, 5].

3 An examination of formal liar sentences

As was shown previously, naive truth theory is considered to be demolished by paradoxes of self-reference, inspiring to create alternative theories avoiding these antinomies. The following approach tries to review the representation technique of the liar sentence challenging naive truth theory.

3.1 Liar sentence translated by biconditional

Many authors discussing the liar paradox represent sentence (2) as was shown in section 2.1, e.g. [2, 4]. Obviously,

$$l \equiv \sim t(\ulcorner l \urcorner) \tag{4}$$

is not a literal translation of (2), thus many authors disagree with this formula.

Another reason against the use of the biconditional here is that according to first-order logic the next logical consequence holds:

$$(p \equiv q) \vdash\dashv \sim(p \equiv \sim q) \tag{5}$$

Interpreting this for (1) yields

$$x \equiv t(\ulcorner x \urcorner) \vdash \sim (x \equiv \sim t(\ulcorner x \urcorner)) \tag{6}$$

that makes the defining scheme of the liar sentence *false* immediately. In other words this translation of the liar sentence cannot cause any paradox relative to truth definition.

3.2 Liar sentence translated by equation

The previous, (4), interpretation of the liar sentence (2) is rather controversial. Another common representation of (2) is the following:

$$\ulcorner l \urcorner = \ulcorner {\sim} t(\ulcorner l \urcorner) \urcorner \tag{7}$$

The only difference between (7) and (4) is that the biconditional is replaced by an equation. This change makes the translation of (2) clearer, although one may still have doubts concerning it, as the consequences remain the same.

At this point contradiction can be deduced, if $\ulcorner l \urcorner$ is substituted into (1), then Leibniz's rule and transitivity of biconditional is applied:

$$t(\ulcorner l \urcorner) \equiv l \equiv {\sim} t(\ulcorner l \urcorner) \tag{8}$$

According to the accepted reasoning this step concludes the commonly known paradox.

Nevertheless, the above deduction suffers from the same defects that in the previous section. It is easy to show by the truth definition (1) together with Leibniz's rule, $(a = b) \supset (F(a) \supset F(b))$, that

$$(\ulcorner x \urcorner = \ulcorner y \urcorner) \supset (x \equiv y) \tag{9}$$

from which by the substitution ${\sim} t(\ulcorner x \urcorner)/y$ and applying the contraposition rule

$$\sim (x \equiv {\sim} t(\ulcorner x \urcorner)) \supset \sim (\ulcorner x \urcorner = \ulcorner {\sim} t(\ulcorner x \urcorner) \urcorner) \tag{10}$$

is reached. Then from this latter and (6)

$$\sim (\ulcorner x \urcorner = \ulcorner {\sim} t(\ulcorner x \urcorner) \urcorner) \tag{11}$$

can be deduced, that contradicts to the assumption (7).

Now it is shown that even the modified representation of the liar sentence, (7), cannot cause paradox relative to the truth definition (1), as the scheme $\ulcorner x \urcorner = \ulcorner {\sim} t(\ulcorner x \urcorner) \urcorner$ is also evaluated *false.*

Essentially, this means that translations of (2), i.e. (4) and (7) hitherto discussed, are unsound as definitions, and thus fail to represent (2) adequately.

4 Finding a way out

Section 3 has shown that the root problem with liar type self-referencing is at an earlier level than was expected. The formula scheme working as the definition of a liar sentence is unravelled as a *false* scheme. This fact seems to weaken the commonly known formal proofs of liar paradox, hence an apparent question probing the grounds of this indefinite phenomenon arises. On the other hand, this antinomy is still present in natural languages [8]. Thus, it cannot be due to a defect in natural language. There is no choice but to assume that the representation itself is invalid.

The question remains open whether a logical type of representation which is able to render this content adequately can exist. The rest of this paper argues that an isomorphic representation is necessary for this purpose, as described in the next section.

5 Representing a liar by iCTRL

The previous sections have shown that traditional language of logic has difficulties in representing self-referential sentences in their natural form. The imperfect fidelity of translation may be a sufficient cause of improper interpretation of the case. Thus, changing the logical representation language may effect a more adequate model of the phenomenon. We present an alternative and novel manner of representing the matter relying on intensional conformal text representation language (iCTRL) initiated by one of the authors [6, 7]. It is a knowledge representation tool closer to natural languages, preserving not only truth as traditional logical language does, but it also models natural grammatical relations. Accordingly, it seems to be suitable for a better formalisation of self-reference, at the same time it allows verification of soundness of the preceding issues.

There is only limited space here to give a detailed formal introduction to iCTRL, so the reader is referred to [6, 7]. Although, a rather reduced sublanguage of iCTRL is sufficient for the whole description, the reader should be made familiar with some new notation. We focus immediately on truth definition. The previously discussed definition (1) can be now written as

$$t\ x, \langle\alpha\rangle x. \equiv \alpha \tag{12}$$

where α represents any arbitrary proposition, $\langle\alpha\rangle x$ the name of this proposition, α, while $t\ x$ stands for the truth predicate, and $t\ x, \langle\alpha\rangle x.$ expresses the statement that α is a true statement.

This latter expression deserves more attention. The predicate symbol $t\ x$ itself appears to be the same as in classical logic, but its application to the corresponding name symbol differs from the way as it is commonly treated: $t(\langle\alpha\rangle)$, since name symbols are treated now as singular predicate symbols, i.e. they must have an argument. Finally "," and "." are punctuation symbols, the former links the predicate of the formula with its subject, providing they share a common variable symbol, the latter closes the formula. Evaluation of the actual sentence $t\ x, \langle\alpha\rangle x.$ simply answers the expectations: it is true if and only if the extension of $\langle\alpha\rangle x$ is completely included in the extension of $t\ x$.

Before trying to represent self-reference, the ordinary case should be presented briefly. Regularly, subjects, like $\langle\alpha\rangle x$, and corresponding predicates, such as $t\ x$, share the same variable symbol, because they are referring to the same group of individuals. However, an exterior subject, mentioned earlier or later in a text, cannot be referred to in this way, only by a compound reference variable term. E.g. $y : x$ can provide that compound term redirecting its left side variable parameter to its right side one that refers to that subject expression sharing the same variable parameter name. Let the pair of sentences *John walks. He whistles.* be considered. The corresponding pair of formulae in this context is $walk\ x, John\ x.\ whistle\ y : x.$ A referred subject can naturally be eliminated by inserting it as it is referred to: $walk\ x, John\ x.\ whistle\ y, John\ y.$[3]

[3] As a matter of fact, subject reference needs a bit more complex notation, which was simplified here to reduce unacquainted formalism to the minimum.

After these preliminary notes the iCTRL formula which exactly formalises the liar sentence (2) is

$$\langle \sim t\ x : y \rangle y. \tag{13}$$

Now the sentence itself plays the role of the subject of the predicate $t\ x$, the variable parameter of which x is redirected to the referred subject by $t\ x : y$.

Considering (12), it does not appear significantly differ from the earlier version of truth definition. Nevertheless, the liar sentence representative (13) is quite dissimilar to the classical formulae, (4) and (7), respectively. According to the construction, it is a literal translation of (2). It does not comprise any extraneous constituents such as biconditional or equation, which are also auxiliary tools in the corresponding classical formulae.

Contradiction results from a substitution of $\langle \sim t\ x : y \rangle y.$ into truth definition (12), that is $\sim t\ x, \langle \alpha \rangle x. \equiv \sim \alpha$ generating:

$$\sim \langle \sim t\ x : y \rangle y. \tag{14}$$

Accomplishment of this substitution appears to be strange, because the liar sentence predicate $\sim t\ x : y$ inside, wrapped into the subject part of the sentence, is to be matched by the left side of (12), $\sim t\ x, \langle \alpha \rangle x$.

The other pair of contradictory statements, similar to (14) and (13), correspondingly causes a paradox. If negation of liar $\sim \langle \sim t\ x : y \rangle y.$, is substituted into (12) that is $\sim (\sim t\ x, \langle \alpha \rangle x). \equiv \alpha$, that similarly yields $\langle \sim t\ x : y \rangle y.$, then that is a contradiction.

In conclusion, the liar sentence has been proven plainly to be antinomic showing that inconsistency based on this kind of argumentation is clearly achieved. Traditional attempts to explain this make the impression that natural language and formal language of logic have drifted apart. However, this is not the case for iCTRL.

6 Closing Remarks

The source of Tarski's semantical paradox has been revised in this paper concluding with the recognition that a liar sentence, which is traditionally applied to generate an explicit antinomy with the classical truth definition, fails to give an effective argument against the related conventional extension of first-order logic. This conclusion can be deduced in each formalisation instance of the classical language of logic originating from the fact that the liar sentence definition fails to define the liar sentence itself. iCTRL modelling enabling a formal syntactic fidelity of translation from natural languages, can prove this paradox case exactly. The approach presented here has shown an adequate representation of self-reference that may stimulate further development with respect to representation techniques of introspection.

Acknowledgements

The core idea of this paper emerged during a colloquium with Prof. Weissmahr. The authors would like to express their indebtedness to him for his invaluable help and encouragement without which this work could never have come to fruition.

The authors would like to thank the anonymous reviewers for their help that made this paper more concise and distinct.

References

1. Gilmore, P. C.: The consistency of partial set theory without extensionality, *Proc. of Symposia in Pure Mathematics*, **13/II** (1974) 147–153.
2. Feferman, S.: Toward useful type-free theories. I., *J of Symbolic Logic*, **49/1** (1984) 75–111.
3. Kerber, M: On Knowledge, Strings and Paradoxes, Proceedings of JELIA, Springer Verlag, 1998.
4. Perlis, D.: Languages with self-reference I: foundations, *Artificial Intelligence*, **25** (1985) 301–322.
5. Perlis, D.: Meta in Logic, in Maes, P.; Nardi, D. eds.: Meta-level architectures and reflection, Elsevier Science B.V., 1988, 37–49.
6. Rédey G.: Conformal Text Representation, *Engineering Applications of AI*, **6** (1993) 65–71.
7. Rédey G.: iCTRL: Intensional Conformal Text Representation Language, *Artificial Intelligence*, **109/1-2** (1999) 33–70.
8. Sainsbury, R.M.: Paradoxes, second edition, Cambridge University Press, 1995.

This article was processed using the LaTeX macro package with LLNCS style

A Critical Analysis of Husserlian Phenomenology in IS Research and IS Development: The Case of Soft Systems Methodology

Stephen K. Probert[1]

[1]Department of Information Systems, Cranfield University, RMCS Shrivenham, Swindon, SN6 8LA, U.K.
s.k.probert@.cranfield.ac.uk

Abstract. Husserlian phenomenology has been used to provide (or perhaps "evoke") the rationale for the use of soft systems approaches in both information systems (IS) research and IS development. The purpose of this brief paper is to encourage a debate about the feasibility and coherence of such projects. A (fairly) typical "interpretive" IS methodology (Soft Systems Methodology) is critically analysed using Adorno's epistemological research as critical theory reference material [5]. It is concluded that the Adorno's arguments against phenomenology apply in full force to IS research and IS development methods (and / or methodologies). Some practical guidelines, for avoiding the problems discussed, will be provided. Finally, some suggestions for further research are given.

1 Introduction

Soft Systems Methodology (SSM) is now often used for both IS research and IS development [1], [2], [3], [4]. Edmund Husserl (1859-1938) developed *phenomenology*, which supposedly provides the foundation of SSM's *epistemology*.[1] Consequently, the arguments put forward in this work will be based (in part) on Adorno's [5] critical analysis of Husserl's phenomenological works [6], [7], [8], [9]. An examination of the reasoning behind the relevant aspects of the SSM advocates' stated (epistemic) position will also be undertaken. The SSM advocates' motivations for adopting the position that they hold will be characterised as the result of a perceived need to attain epistemic certitude. Further considerations concerning Husserl's search for certitude will inform the pragmatic recommendations, which will be discussed at this point. The practical conclusion drawn will be that, whilst not attempting to sanction "sloppy" systems analysis, epistemic certitude is not attainable

1 "Soft systems methodology implies ... a model of social reality such as is found in the ... (phenomenological) tradition deriving sociologically from Weber and philosophically from Husserl." [1] (p. 19).

- therefore the demand for it can only be counter-productive – when undertaking IS research and (*a fortiori*) IS development.

2 Imaginary Altitude

Essentially, SSM advocates hold that statements about the real (i.e. objective) states of affairs in the social world are unwarranted and untenable. Consequently, discourse about mental states is *elevated* to a position of high (or higher) epistemic significance and statements about the real world are denigrated as having a low - or even insignificant – epistemic status [10], [11]. It is precisely this *elevation* (of discourse about mental / ideal states of affairs) which constitutes the common ground between the SSM advocates and Husserl; this (generic) approach is criticised by Adorno for what he characterises as its *imaginary altitude. Prima facie* the (crude) positivists' position is that sense-data puts us in immediate contact with external reality (although considerable variations on this theme can be found in the writings of the so-called positivists). At any rate, it is this (a somewhat "straw man") version of positivist thought that both Husserl and the SSM advocates take umbrage at. The SSM advocates have often proffered the view that, as ideal "mental constructs", *human activity systems* (i.e., in their view, *information systems)* have properties, characteristics, etc. which may be examined; whereas – on the contrary – *human activity systems* as real world occurrences are strictly-speaking unknowable and therefore they cannot be modelled. In this respect Husserl's arguments and those of the SSM advocates (for subjective idealism) are strikingly similar. Adorno argues that the motivation for idealism lies within the belief that unless a thought (or a judgement) about some aspect of experience admits the possibility of being *certain* (whether true or false) then that thought is epistemically worthless, "The thesis of the perceptibility of the purely possible as a doctrine of essential insight, or as Husserl originally called it, categorial intuition, has become the motto of all philosophical approaches which evoke phenomenology. The fact that the new method should guarantee ideal states of affairs the same immediacy and infallibility as sense-data in the received ["positivist"] view, explains the influence which Husserl exercised over those who could no longer be satisfied with neo-Kantian systems and yet were unwilling to blindly hand themselves over to irrationalism." [5] (p. 200).

The "altitude" supposedly gained by taking such a view (i.e. the idealism adhered to by both the SSM advocates and Husserl) is achieved by, as it were, "rising above" the real world into an ideal world (or worlds) – in a search for greater epistemic security. Of course, the "price to be paid" is in the removal ("elevation") of oneself from the real world within which one may be attempting to act. However, and in agreement with Adorno, it is not being suggested here that an alternative position of naïve positivism should be adopted, as "[C]ategorial intuition is the paradoxical apex of his [Husserl's] thought. It is the indifference into which the positivistic motif of intuitability and the rationalistic one of being-in-itself of ideal-states-of-affairs should be sublated. The movement of Husserlian thought could not tarry at this apex. Categorial intuition is no newly discovered principle of philosophizing. It proves to be a sheer dialectical moment of transition: imaginary altitude." [5] (p. 201). However, it might be argued that the

SSM advocates in fact hold the position that thought is not *so* detached from the real world as my characterisation (above) would imply. Indeed, the SSM texts contain many references to an unfolding flux of ideas *and* events. However, it is also made clear – in the various SSM texts (e.g. [1], [2], [4], [10], [11]) that "perceived events" are just (precisely) *subjective perceptions*. Adorno cogently distinguishes between epistemological accounts of experience given in terms of *sense-data of* and (ephemeral) *encounters with* the real world, "In a certain way categorial intuition was devised by the doctrine of propositions in themselves … If these are truly to be more than creations of thought, then they cannot really be products of thought but must simply be encountered … by it. The paradoxical demand for a merely encountering thought arises from the claim to validity on the part of logical absolutism[2]. The doctrine of categorial intuition is the result of this on the subject side." [5] (pp. 201-202).

3 Subjective Certitude and Epistemological Rigour

The question that will now be asked is: why should so much emphasis be placed on (the possibility of) subjective certitude in such formulations of the epistemological problem of knowledge-discovery? According to Adorno, the answer is to be found in the actual circumstances in which academics find themselves, i.e. middlemen. Interestingly, this argument would appear to hold a fortiori for the likes of SSM practitioners, consultants, etc. The demand for subjective certitude – inherent in the epistemology proffered by the SSM advocates – would prima facie seem to generate immediate problems for the use of SSM's epistemology in practical endeavours. One might think that practical IS development work should – minimally – be more concerned with getting a practical working knowledge of a situation in order to take positive action – rather than getting embroiled in "epistemologically purist" issues and concerns. Of course, to take this literally would be to proceed uncritically. In order to operate in a critically aware manner, epistemological considerations will be important – but it is argued here that "epistemological purism" is not the best way to proceed. Further discussion of an appropriate epistemological framework with which to undertake critical systems analysis lies outside the scope of this paper (which is not to suggest that it is unimportant).

At any rate the source of the subjective idealism - inherent in the SSM advocates epistemological accounts – may be found in practice rather than in theory. The accounts of epistemology given in the SSM texts are supposedly based on (or supported by) the practical experiences of using systems ideas in organisations. In all such accounts (encountered by the author at any rate), the Soft Systems Practitioner does not claim to be the owner of the system. Indeed, the impression one gets is usually of the SSM practitioner being rather unceremoniously "dumped" into a conflict-ridden and potentially hostile social situation of which he or she has little prior knowledge. From a

---

[2] The term 'logical absolutism' is introduced by Adorno to connote Husserl's view of logical statements as being in no way dependent on events occurring in the real world for their truth-values; this is an important aspect of Husserl's conception of eidetic sciences.

critical perspective, this is significant, because Adorno's critique of Husserlian phenomenology does not depend on (the success of) a purely rational critical exercise. Rather, it depends on the reinterpretation of epistemological categories as the *products* of social conditions – particularly those where power / violence / intimidation / etc. is exercised. Jarvis explains Adorno's *metacritique* project thus, "It was in the work towards the Husserl book [5] ... that the mature form of Adorno's thought began decisively to emerge. This was Adorno's most extensive attempt to date to justify in detail his belief that even those philosophical texts which were apparently most abstract necessarily contained sedimented within them the traces of the social experience which had made them possible.. In particular, it is in this study that Adorno begins to put into practice the idea of a *metacritique*. Whereas epistemological critique asks what experiences make experience possible, metacritique asks what experiences make the epistemological categories possible." [12] (p. 12). Might the social conditions (alluded to above) explain the perceived need for, or the motivation for seeking, certitude? Adorno makes the following comments about subjective idealists (in general) in the introduction to his *Against Epistemology – A Metacritique*[3][5], "The open or secret pomp and the totally unobvious need for absolute spiritual security – for why, indeed, should the playful luck of spirit be diminished by the risk of error? – are the reflex to real powerlessness and insecurity. They are the self-deafening roar through positivity of those who neither contribute to the real reproduction of life nor actually participate in its real mastery. As middlemen, they only commend and sell to the master his means of lordship, spirit objectified ... into method [or methodology, for that matter]... They use their subjectivity to subtract the subject from truth and their idea of objectivity is as a residue." [5] (p. 15). The practical problems generated in IS research and IS development are unlikely to be solved by the adoption of an impractical epistemological standpoint.

4 Conclusion

It is concluded that – whatever the motivations for desiring it – epistemic certitude is not attainable - therefore the demand for it can only be counter-productive – when undertaking IS research and IS development. Our understanding of the real world in which IS research and IS development must take place may often be partial, confused and even bigoted. Essentially, critically-minded vigilance will provide some defence against the latter – as will an openness to the critical comments and suggestions of others. For the former – the epistemological problems – we had best learn to make do with whatever understanding of the problem situation can be obtained, given the time and resources available. This is *not* to sanction sloppy analysis! The alternative - only to sanction (unattainable) epistemological rigour - can only force us to withdraw our attention from the real world and into our (subjective) selves. Few practical problems

[3] The title of this book is somewhat misleading, as – in it - Adorno is conducting a critical analysis of subjective idealist epistemology as a (sort of) groundwork for an alternative epistemology, "Criticizing epistemology also means ... retaining it." [5] (p. 27 [N.B. the three dots are included in the original text]). Some aspects of what such an alternative epistemology might look like are discussed by Guzzoni [13].

are amenable to solution solely by introspection - although this is not to deny the value and importance of critical reflection. It should be noted that Husserl's epistemological arguments, and Adorno's critique of these arguments, are both extremely difficult topics; further research is warranted here. At any rate, the uses (or abuses!) of such complex arguments to legitimate approaches to IS research or IS development are fraught with difficulties – especially when the practical consequences of these approaches have not been adequately considered. There is a need for more *critical* research here also, including *metacritical* research – as has been carried out above.

References

1. Checkland, P. B.: Systems Thinking, Systems Practice. Wiley, Chichester (1981)
2. Checkland, P. and Scholes, J.: Soft Systems Methodology in Action. Wiley, Chichester (1990)
3. Lewis, P.: Information-Systems Development. Pitman, London (1994)
4. Checkland, P. and Holwell, S.: Information, Systems, and Information Systems – Making Sense of the Field. Wiley, Chichester (1998)
5. Adorno, T. W.: Against Epistemology: A Metacritique. Basil Blackwell, Oxford (1982)
6. Husserl, E.: Ideas. George Allen and Unwin, London (1931)
7. Husserl, E.: Logical Investigations. Routledge and Kegan Paul, London (1970)
8. Husserl, E.: Phenomenological Psychology. Martinus Nijhoff, The Hague (1977)
9. Husserl, E.: The Idea of Phenomenology. Kluwer Academic Publishers, Dordrecht (1990)
10. Checkland, P. B.: Towards the Coherent Expression of Systems Ideas. Journal of Applied Systems Analysis 18 (1991) 25-28
11. Checkland, P. B.: Systems and Scholarship: The Need to Do Better. Journal of the Operational Research Society 43 (1992) 1023-1030
12. Jarvis, S.: Adorno: A Critical Introduction. Polity Press, Cambridge (1998)
13. Guzzoni, U.: Reason – A Different Reason – Something Different Than Reason? Wondering about the Concept of a Different Reason in Adorno, Lyotard, and Sloterdijk. In: Pensky, M. (ed.): The Actuality of Adorno. State University of New York Press, Albany (1997) 23-42

Is Bayesian Inversion a Model for Searching the Truth?

Wolfgang Jacoby, Mainz
jacoby@uni-mainz.de

The philosophical approach to searching the truth is akin to geophysical Baysian inversion of potential fields. Karl Popper proposed that science approaches the infinitely distant truth by challenging and falsifying existing hypotheses and theories. The kinship between the two is that "truth" is as much hidden in both fields of human endeavour.
Bayesian inversion is built on the assumption that a "true" model exists which gives observable signals which are used to determine and to optimize the model parameters. For this we must know the "sensitivities" of the observations to changes of all the adjustable model parameters (i.e. the Jacobian). In geophysical potential fields as gravity and magnetics, inversion is infinitely ambiguous. An infinite set of mass models exist which generate the same external field; in other words, an infinite set of "difference models" generate no external field, the set is called the "null space". On the other hand, there is an even larger model set which generates different fields. Any of such models can be excluded by gravity inversion.
A reduction of the null space is possible by invoking a priori knowledge. It is an indispensable precondition for any meaningful inversion to be obtained. The Bayesian approach is to regard the a priori information as having the same categorical status as the observations to be "explained" or "fitted". Both are treated in the same way. A priori model parameters are "optimized" within their error limits, as the model effects are fitted to the observations within their error limits. Solutions are as reliable, as the errors or uncertainties allow. It is therefore essential to estimate the errors and model uncertainties as "carefully" as possible.
A few examples are presented to make the point.
So, I think that philosophy is an inversion procedure.
Philosophically, we attempt to understand complex and multiple observations and confront our pre-existing ideas, hypotheses, theories with them. We are in the same situation as the geophysicist. We build a models and "calculate" or predict their effects or consequences which can be checked by observations. From Bayesian inversion we can learn that we must be most concerned about the uncertainties of any of the complex model features and predictions. We must also be fully aware of the fact that agreement between prediction and observations is only a necessary condition, not a sufficient one. As in gravity, probably an infinite set of hypotheses exist that may predict the same or similar observations or experiences. As in gravity, this problem may be minimized with a maximum of additional information. For philosophy it suggests that complex "interdisciplinary" sets of ideas, models and observations must be combined. But do we generally know the uncertainties of our data, hypotheses, ideas? And can we claim to approach the truth in any quantifiable way? Do we know the sensitivities of our theory parameters to observable phenomena?

Author Index

Proceedings of the First Workshop on Philosophy and Informatics WSPI 2004, Cologne, Germany

RR-04-02

Research Report