Variations and Application Conditions Of the Data Type »Image«

The Foundation of Computational Visualistics

H a b i l i t a t i o n s s c h r i f t

zur Erlangung der *Venia legendi* für

Computervisualistik (computational visualistics)

angenommen durch die Fakultät für Informatik
der Otto-von-Guericke-Universität Magdeburg

von: *Dr. rer. nat. Jörg R.J. Schirra (Dipl.-Inform.)*

geb. am 03. August 1960 in Illingen (Saarland)

Gutachter:

Prof. Dr. Thomas Strothotte
Prof. Dr. Wolfgang Wahlster
Prof. Dr. Arno Ros
Prof. Dr. Jerome A. Feldman

Magdeburg, den 11. Mai 2005

A NOTE AHEAD

This book is a map. It maps the landscapes of the country of digital images, or, as it was lately renamed, the realms of Computational Visualistics.

Like any picture, a map – and hence this book – is a context builder: it allows the readers to explore different paths in an abstract region, to connect many landmarks on several ways, and to establish their own distinctions of figures and backgrounds according to their proper interests.

However, a text is bound to its linear progression of propositions woven into the digital fabric of argumentation that only mimics the true spatial quality of images. As an extended path, reading this text snakes through the map in the effort to systematically cover all of its regions: the map only appears in the reader's mind. Not all of the details present may be integrated on first view. After all: a real map presents all its details simultaneously, but only those details are actually "read" that are relevant for the reader's present intentions. The map reveals its contents not on a single glance. That is to say: this text is explicitly written in order to be read more than one time.

Concerning the sources of the ideas described: it is a valuable academic tradition to mention all means used, well observed by the bibliography at the end of this book. There is however a problem in the strict application of this principle depending on the enormously extended medial access to the thoughts of others. It has actually become impossible to explicitly quote or even remember everything that has contributed ideas to an ambitious academic work: apart from the classical form of scientific papers, books, talks, discussions, dialogues, and (long ago) lectures, which usually can be traced back easily, there have been documentaries in television, features in radio, articles in newspapers, fictional films and novels, comics and advertisements transmitting views effective in this book; from visits to exhibitions to web-browsing, many other kinds of mediated communication have provided arguments to the present text without the author being able to remember them in detail.

Although I am not able to trace back all the "underground" elements not originated by the author, and to list their sources: without them, this work would not have been possible.

Magdeburg, October 2003

Table of Content

1 Images in Computer Science: Clarifications Required

1.1 The Age of the Images

Images take a rather prominent place in contemporary life in the western societies. Together with language, they have been connected to human culture from the very beginning. Recently – that is, after several millennia of written word's dominance – their part is increasing again remarkably. Can't we even characterize the 20th century as the century of pictures? Photography and film have reached a heyday barely anticipated when they were invented at the end of the 19th century. Together with TV and video, they have become generally accessible and easily consumable pictorial media, which partially can even be produced by everybody without many problems. FAX and Xerox copies allow us for about 40 years now to get in the most simple way copies of graphics, and to transfer them almost immediately to the most distant places. Comics and tabloids with many photographs have used the new technologies of picture production to renew and multiply a tradition that reaches from the Neolithic paintings through the Bayeux tapistery to WILLIAM HOGARTH, RODOLPHE TÖPFFER, WILHELM BUSCH, and further on. The effects of the digital revolution during the last three decades on producing, distributing, and "consuming" pictures are yet hardly conceivable in their totality. This is true not only for the entertainment industry, which has developed into a significant factor of economy already (concerning its commercial weight alone). In the area of education, the importance of supporting learning with modern pictorial media is basically unquestioned, as well. Even in scientific discourse, graphical representations have become unavoidable – in didactical contexts as in diagnostic ones: otherwise the growing complexity of research themes cannot be presented in an adequate manner that is simultaneously accessible fast enough. In general, skilled work without using pictures by means of computers is receding quickly: we barely can imagine our society without the graphic programmes for designers, the ultrasonic diagnostic units for physicians or the digital simulation models for engineers.

The fact of a waxing "pictorialization" of our environment, be it private or at work, has been judged quite antithetically [POSTMAN *1985*]: on the one hand, images are ascribed the potential to let us gain a fast and trustworthy orientation about complex matters. Digital pictures in particular open us new ways for accessing reality, and help to make traditional (i.e., mostly verbal) approaches more easily accessible. On the other hand, critical minds deplore the erosion of rational structures of discourse and thought associated with the flood of images: the medium of written language alone, they state, supports and advances a conceptual discussion of reality and knowledge. Indeed, the antagonists of this strange discrepancy consider rather different phenomena by the expression "image": the first value special aspects of modern technology while the second judge structural implications of modern entertainment industry.

Steps toward a general science of images, which we may call "general visualistics" in analogy to general linguistics, have been taken recently. So far, a unique scientific basis for circumscribing and describing the heterogeneous phenomenon "image" in an interpersonally verifiable manner has still been missing while distinct aspects falling in the domain of visualistics have predominantly been dealt with in several other disciplines – partially even the same aspects in incompatible manners. History of arts and aesthetics, philosophy and semiotics are traditionally involved. Psychology and science of commu-

nication, anthropology and science of media have joint more recently. Last (though not least), important contributions to certain aspects of a new science of images have come from computer science.

1.2 Toward Information Society

Picture's triumph in the 20th century is embedded in a general tendency of alteration in western society: numerous analysts tell us that we change rapidly into an "information society" where most of the labour is gathering or processing information instead of material goods. Pictorial information here plays a prominent part. Consider, for example, the enormous amount of earth-related information gathered by satellites day after day, and the important, sometimes even explosive political and economical effects, an appropriate processing and presentation of that information may provoke. Who does not remember the impact of reports on the "ozone hole" gained in particular by the visualizations: seemingly, mere columns of figures could not have informed us in a sufficiently sensible manner about the expansion and temporal development of damage in the Antarctic region.

As most of the information-related work characterizing information society is performed by means of technical tools, the expression 'media society' is used, as well. The appearance of new keywords like 'multimedia', 'internet', 'information technology' (now even commonly shortened to 'IT') bears witness of the continuous social changes that are transforming every single aspect of society – public or private; economic, cultural or scientific. On the way toward information society, the forms of communication in particular are altered as their characters depend on the media used. The concept »information« is often determined as: "a mediated message with pertinent meaning for sender and receiver" (cf., e.g., [PROSS *1972*]). Correspondingly, the expression 'medium' is used in general to indicate a means for transferring and distributing information – the "middle area" between sender and receiver in a common spatial metaphor of communication; an "in between" that is simultaneously connecting and separating the communicative partners. Its structure determines the form of messages possible to pass. A well-known classification system of media theory distinguishes three types of media: whereas media of class *I* (also called primary media) do not involve any technical devices that open the possibility of temporally or spatially separating the communicative partners, class *II* media (secondary media), like books or letters, involve devices on the producers side, and class *III* media (tertiary media) on both sides of the communication channel, like TV or telephone. With the shift toward information society, class *III* media are becoming the dominant means of exchanging information.

While speaking of information implies communication, i.e., some interaction between several partners, 'data' and 'knowledge' – two expressions sometimes used almost synonymous with 'information' – lack such implications: data is (potential) information considered from a merely technical point of view, e.g., the data of ozone concentration in the stratosphere gathered by satellites. The expression 'knowledge' comes into the game when information is involved in the (conscious) decisions of somebody to act in a particular way (or the explanation thereof), e.g., the knowledge about the ozone hole influencing a citizen's political decision. Thus, in another sense of mediation, information and its form has to be conceived of as mediating between data and knowledge.

On the long turn, the construction of fictional visual presentations up to "virtual realities" as they are commonly known may have even deeper social consequences. Hollywood film productions provide a number of quite prominent examples for the potential

of such electronic picture production and manipulation. The economic potential of visually intriguing computer games cannot be overestimated while their other effects on society are still to be investigated further. With photography and cinema, the technological ascent of pictures gains its modern dynamic, but only in its younger digital form, manipulating graphical information has become possible almost without any limitation. In the entertainment industry, such "makeovers" can be quite desirable; in political and economical contexts, however, catastrophic consequences are imminent.[1]

It is, of course, computers that have contributed most to accelerate our transformation to information society in the last couple of years, and that have empowered us to manipulate pictures almost beyond imagination.

1.3 Images and Computers: The Digital Picture

In fact, the combination of images and computers did originally cost the former a property conceived of as characterictic for pictures by the scientists of many disciplines involved: pictures had to become digital in order to join that liaison. Essentially this means that the resolution of pictures has a definite (and often quite small) value. In contrast, the common view holds that pictures have to be (at least in principle) analogous, i.e., without any limitation of resolution. This debate is still a theoretical issue we shall discuss in greater detail below, but for pratical reasons, the restriction is quite irrelevant as the resolution can be chosen far below the threshold of our visual resolution. Far more relevant is, however, the question of authenticity for images being digitally processed, and also the question of their communicative and expressive forces.

The rapid alterations into an information society, which sometimes are even equaled to such major leaps in human development as the Neolithic revolution or the Industrial revolution, have provided pictures with a particular feature they have rarely shown so far: interactivity, i.e., the potential to be modified instantaneously by the beholder. Those alterations may be concerned mainly with parameters of the screen, but also with attributes of the scene depicted (including the beholder's relative viewing position). In the latter case, we reach the fascinating field of 3D virtual environments, more popularly known as virtual reality. It is in fact an open question, whether these systems are to be conceived of as pictures or rather as an architecture or sculpture.

In the following, we shall use the artificial expression "computational visualistics" for addressing the whole range of investigating scientifically pictures "in" the computer. The expression was first used in 1996 for an academic educational programme, mirroring the relation to computational linguistics – the field of investigation concerned with (natural) languages "in" computers. In a way, this book is essentially about the question whether computational visualistics can be constructed as a homogenous field of research (in contrast to a mere agglomeration of several picture-related areas of computer science). In order to positively answer that question a unique subject has to be specified together with a particular methodology. For short: The subject of computational visualistics may best be described as the data type »image« and its implementations. Its methodology is essentially derived from computer science with an interdisciplinary component from the general science of pictures. We shall come back to those questions in much greater detail below.

[1] The observation that the public information in recent wars has predominantly been made of digitized pictures may be mentioned here only as a secondary thought.

The theme 'pictures and computers' indeed forms an extremely manifold domain, which is often quite hard to follow up in its complexity and variety. But it also offers insight with unforeseen potential and risk for future development. Computer scientists must react upon this challenge in a manner adequate to the most developed social standards of our time. That is, not only do they have to keep up to date with objective knowledge and skill, but also with respect to the understanding of their social functions and tasks.

1.4 Requirements for a Modern Computer Scientist

C.P. SNOW's diagnosis that modern (Western) society is split into "two cultures" still holds after 40 years [SNOW *1959*]. In his Rede Lecture, SNOW used that expression for critically referring to the communicational breakdown between art and the humanities on the one side, and science and engineering on the other side. Since recent developments in teaching have to be seen in that light, a closer look at the underlying difference may help to better understand the conception of "new" engineers.

In the nutshell, engineering is the endeavor of constructing systematically material artifacts – engines – that are defined by some given purpose: if they serve that purpose they "work", if they don't work they are "broken". This type of activity can be understood as one of the most prominent consequences of a shift in the late medieval period, prepared by BACON, explicitly stated by GALILEI, and made an ideology by DESCARTES [ROS *1990, Vol. 2*]: a shift that broke loose the enormous acceleration of the technical development of the following four centuries. This was to start focusing more or less exclusively on how nature can be used for our goals as the only guiding principle for rationality of arguments. The ancient philosophers had sought to understand nature in its own right, without projecting our own views. However, this became problematic, since an access to the nature of things could not be rationally defended. Understanding nature seemed possible only as a means of dominating nature. Engineering comes, so to speak, as a late consequence of the biblical "subdue the earth" (Gen. 1.28).

The humanities, on the other hand, are usually conceived as an investigation following the old Delphian motto "gnothi se auton", "know thyself": the unremitting endeavor of self-interpretation, where human beings try to understand their very own nature. The roots of dealing in a systematic manner with the questions of self-knowledge, which also include the ethical component "How do we want to live?", stems from the ancient Greek philosophers about two and a half millennia back. Human beings, as the central object of investigation, are conceived of as ultimately setting their goals and purposes on their own: unlike a machine, a person not following one's goals is not "broken", but follows his/her own goals. The actions of that person must be rated with respect to the objectives uttered by herself/himself. This also includes the actions of research. The reflexive nature of such a hermeneutic investigation must lead to standards of rationality and methods of argumentation that are rather different from the empirical sciences or engineering [BROOKS *1996*].

Even with this simplified sketch, it is clear that the underlying methodologies of the two "cultures" are quite conflicting: constructing machines that follow some pre-set goals vs. interpreting phenomena related to the self-determined aims of humans. The success of the scientific-technical culture with its strictly purpose-driven arguments certainly speaks for itself. However, the underlying programme of "subdue the earth" is not uncontroversial: who sets the goals pursued? Who decides about the purposes that rule development and application of technologies? The critique of a purely technocratic per-

spective can be heard louder and louder since, at least, the late 1960s. An integration of the two methodologies becomes increasingly necessary if the problems – often evoked by the very use of engineering – are to be solved. In the words of the German philosopher HABERMAS, this is the question of whether our societies are able to find a satisfying relation between our enormously grown technical powers, and democracy as the institutional forum for discussing how we want to live [HABERMAS *1996*].

A general solution of how to integrate the culture of "subdue the earth" with that of "know thyself" cannot be approached here. However, "building bridges" from both sides is a task worth trying. In future, a successful engineer "must, in addition to being competent in engineering, be a skilled listener for concerns of customers or clients, be rigorous in managing commitments and achieving customer or client satisfaction, and be organized for ongoing learning." [DENNING *1992*]. A shift can be observed away from conceiving of engineering as merely "art for art's sake" towards a communicative expertise of assisting other people in solving their particular problems.

1.5 Determining the Goal

In the face of the eminent role of pictures generated, processed, stored, manipulated or transferred by computers in the progress of social regrouping toward information society, it must be considered as crucial for every computer scientist involved to understand the underlying abstract data structure and the reasons for its properties. The question, thus, ultimately is: What are images (and their uses) for computer science, and what is computer science for images (and their uses). The profound understanding of the own position on all levels of the intellectual environment is important for planning successfully any further development: this includes the development of specific technical solutions with computerized uses of images, and, on a more general level, the direction of research leading to completely new technologies.

Furthermore, the ability to clearly lay open the basis of one's own professional decisions is important for the proper external presentation in particular to those that are affected by those decisions.[2] This includes in particular the scientists in the other disciplines of general visualistics using the results of computer visualists. The relation between the decisions in computer science and the arguments structuring the fields of application are obviously highly relevant; but they remain often quite unclear. What is needed is a description and justification for the particular methods and subjects of computational visualistics.

The following argumentations are guided by the idea that only results of general visualistics gained in an interdisciplinary manner provide us with an adequate framework for generating and employing pictures in human-computer interfaces. Which properties and relations are absolutely needed? What ranges of freedom can or must be granted? Which additional parameters may or may not play a role depending on the particular task at hand? The elaboration of those structures has to be conceived of as a sub-domain of general visualistics, i.e., in close relation to its other sub-domains.

There are of course many texts dealing with pictures in computer science from a general perspective. They fall in two classes: one type assumes that the concept »image« is already completely clear – usually employing a rather naïve and narrow understanding, and emphasizing technical aspects of generating or manipulating digitized images.

[2] The expression "collateral damage" may come from a different field; but it evokes quite an adequate image in the context of unreasonably introduced technical artifacts, too.

Members of the other class investigate from a more sociological or media-theoretical perspective the influence computers have on image uses and image users; they are often not interested too much in technical details.

The main goal of the present text is to integrate the two perspectives and to provide a sketch of a general investigation of the concept »image« from the particular point of view of computational visualistics.

To that purpose, chapter 2 starts with a general overview about computational visualistics – and thus, on the approach of computer science to images – by sketching the root disciplines computer science and general visualistics, their subjects, and their methodologies. This leads to the introduction of our main theme, the data type »image«, and a first set of coarse sub-divisions.

Chapter 3 summarizes some of the theoretical approaches to images (and pictures) in other disciplines, i.e., what any computational visualist has to know from other areas in general visualistics. The relations between images and (a) what is depicted, (b) what we can communicate in contrast by means of language, and (c) what image users do in general when communicating are recapitulated on the basis of a definition of pictures as used in visualistics.

On this basis, chapter 4 elaborates the relations and attributes of the generic data structure with the type »image« on a general level. Following the semiotic distinction between syntax, semantics and pragmatics of signs, the relations between several parts of the generic data structure are investigated. This includes in particular: types for geometric Gestalts determining the pictorial syntax; the relation between geometry and sortal objects (spatio-temporal, material, countable entities in the usual sense) as the basis for semantic analyses; and beholder models as the means to deal with pragmatic aspects.

Chapter 5 introduces a collection of four case studies that demonstrate various dimensions of the data type »image« as introduced in the preceding chapter.

Finally, chapter 6 summarizes the whole investigation and presents some perspectives concerning the future development of computational visualistics.

2 Computational Visualistics: Seen From Its Roots

Computational visualistics gains its name from two "parent disciplines": "computational" refers to the rather young discipline of computer science, which nevertheless is well-established for about 30 years at our universities. "Visualistics", on the other hand, brings into mind a unified science of pictures – general visualistics – that has not institutionally existed before recently. A close look on these sciences, their subjects and their methodologies seems prudent in order to gain a better understanding of the scientific basis of computational visualistics, and may additionally provide us with a more precise plan for our investigation.

2.1 Computer Science: Subject and Methodology

Let us focus our attention onto methodology first: computer science, the endeavor of studying scientifically computers and information processing, has two different roots determining its methodology. In some aspects, computer science is a typical *structural science* like mathematics and logic: their subjects are purely abstract entities and their relations – entities far off of our living practice, at best linked to everyday life by means of an interpretation relation. With respect to some other aspects, computer scientists are like electrical engineers interested in *engineering* problems, an interest resulting in concrete artifacts that have already changed our lifes dramatically during the past few decades and continue to do so with growing acceleration. The fluctuation of the focus of attention between structural science and engineering is characteristic for all investigations in computer science, and thus, is valid for the dealing with pictorial data, as well. On the one hand, particular abstract data types for pictorial representations are investigated and designed from a purely structural point of view. For example, efficiency properties are examined, or minimal sub-structures for particular tasks determined. On the other hand, concrete algorithms (based on those data structures) for, e.g., picture processing are "software-engineered" and used in diagnosis – with considerable influence on our social structure.

Correspondingly, computer science's subject is a pair, as well. Although it is not wrong to view computer science as the discipline dealing scientifically with computers and data processing – as we often do colloquially – a better understanding evolves if we consider »data structure« and »implementation« as the basic concepts and main subjects of the field, two concepts that can more easily be related to central concepts in the philosophical theory of argumentation. That relation is particularly helpful to understand the connection between computer science and its application domains.

The processing of data is certainly a crucial theme for computer scientists, but it depends completely on the fact that data is always structured and grouped into types. Each such type implies a set of possibilities to "do something" with that kind of data: numbers can be added or multiplied (etc.); polygons in a geometric model can be moved or turned, mirrored or strained (etc.), but not *vice versa*. Usually, several data types and their interactions are relevant. As it is only important here that we can perform some operations with one sort of data so that certain relations hold between their results while ignoring the concrete manner of how those operations are actually realized, computer scientists consider *abstract data structures* – abstract entities that grasp exactly the essential properties. Algebraic formulae or logical expressions are often used to that purpose: the former for describing which operations transform the instances of which data

type into what other type's instances; and the latter determining which properties remain unchanged – invariant – after a certain sequence of operations [EHRIG & MAHR 1985]. The methodology of computer science as a structural science is, then, partially covered by this first question: How can we find for a given class of problems an adequate data structure so that a procedural solution – an algorithm – can be given by means of combining the operations of that data structure?

A close relationship between abstract data structures and the understanding of a field of concepts can be seen when taking into account the philosophical theory of rational argumentation – an association that is also particularly well suited for studying the relations between computer science and other disciplines. We shall therefore elaborate this unusual approach to the subjects and methodologies of computer science a bit further.

If we refer by the expression 'the concept »*X*«' – e.g., by 'the concept »image«' – to everything that is structurally common to all explanations of '*X*' (in the example: the expression 'image') and its synonyms [WITTGENSTEIN *1953*] – that is, everything that "remains the same independent of how or in what language I formulate or show it" – then naturally, we never examine one concept alone: it is always a system of concepts that are mutually related and cannot be defined independently from each other, like »king«, »queen«, »knight, and »medieval society» (or alternatively »chess«) or, of course, »image« and »perception«. They belong to the same *field of concepts*. From the perspective of structural science, we can therefore view data sorts as a formalized version of certain concepts, and the corresponding data structure as the appropriate field of concepts. While concepts and their fields in everyday life often lack precision or may even be inconsistently organized, abstract data types must (usually) satisfy formal rules of consistency and completeness.

Relations between several fields of concepts are of particular interest for the theory of argumentation. The internal relations of one field may indeed be used to explain correct or wrong applications of the concepts of that field (or the expressions for these concepts) – presupposing however that all the parties involved in the argumentation agree that the field considered is appropriate at all. But in order to firstly motivate this presupposition for a critically-minded interlocutor: in order to explain why the internal rules are adequate conceptual rules in the frame of a rational argumentation, field-external relations have to be thrown into the game, in particular relations to fields of concepts all the parties of the argumentation agree upon already [ROS *1999*]. We may try to reconstruct for our opponent the conceptual structures of the field in question as a systematic combination of the concepts already shared.

Take an example from mathematics: new types of numbers are introduced exactly with such a reconstructing schema. Imagine we only know about integer numbers and are to be introduced to rational numbers. Perhaps, somebody (let us say, a globe trotter interested in mathematics) told us about this – for us new – kind of numbers he heard of in Arabia, and we, on first view, experience the described entities and their properties as rather strange. Or we spontaneously invented the specification (the description of the internal rules) like in a combinatorial game without being aware of doing more than a "*Glasperlenspiel*". In any case: the only thing we know for the time being is the abstract and symbolic specification of that concept. Whether such entities *really* exist, i.e., whether we deal here with a useful and correctly constructed concept, that is still completely unclear.

How could our dialog partner (the mathematical globe trotter) convince us that these mathematical entities, which for us seem so strange, are possible and useful ("real ob-

jects", so to speak)? He could try to show us how to introduce this concept from the fields of concepts we already have (namely integer numbers). That is, he can show us the schema for implementing rational numbers by means of integer numbers (as equivalence sets of pairs of integers, to be precise, i.e., as fractional numbers). That schema must specify how the primitives and the operations of the rational numbers are constructed using the primitives and operations of the integers. This could be done by means of a constructive operation that is neither part of the field of integers nor of the rational numbers. Our teaching dialog partner may postulate, for example, that every instance of a rational number can be represented by pairs of instances of integer numbers: we are very well able to recognize such pairs, following our preconditions. If we accept this introduction schema for the rationals, we are additionally able to *justify* (ground) the internal rules of the new field as given in its specification by means of the attributes of integers: that the equivalence class x/x (for all integers x that are not 0) is the neutral element of rational multiplication can now be derived from the rules of the integers and the fractional combination schema (etc.).

Analogously in computer science, an abstract data structure can be *implemented* by means of other data structures: the implementation provides us with "real" instances of data types that had only been symbolically defined by means of the abstract descriptions of the data types included. Furthermore, a computer scientist may motivate that an abstract data structure (and a particular algorithm defined within) does indeed "make sense" (i.e., does what we want it to do): he may do so – in a scientific paper or talk, for example – by pointing out the construction schema of the data structure by means of those data structures supposedly accepted by her audience in advance, i.e., by giving a corresponding implementation.

Thus, »implementation« is a central concept of computer science derived from the notion of data processing. But it is also closely linked to computers, the second subject of computer science in the colloquial understanding: for the engineering perspective, computers are in fact implementation engines. If, for example, a group of engineers has reached an agreement that a certain artifact of electrical engineering indeed realizes the data structure of the integer numbers – i.e., the artifact "acts" like that (at least if no technical error occurs) – then, of course, the engineers can perform particular calculations with integer numbers by means of the artifact. But they may also use several copies of the artifact for constructing another technical artifact – an artifact they are motivated to view as a realization of another data structure, e.g., the rational numbers, if its construction mirrors the abstract implementation schema of that data structure on the basis of the integer numbers. Therefore, realizations of an abstract implementation schema are often called "technical implementation" (or "implementation in the technical sense"). The engineers may use the new artifact for doing calculations with rational numbers. But they may also convince other persons (who agree already on the interpretation of the "integer artifacts") of that understanding of their "rational number machine" by explaining the abstract implementation schema.

Computers are a particular sort of engineering artifacts that – by general understanding – provide through a chain of realizations of more elementary structures (e.g., assembler and register machines, binary numbers and logical gates, electron flows and magnetic bubbles, to name but a few) a technical implementation of a broad spectrum of useful data structures chosen in a way that one can use them to implement more or less

easily any other data structures.[3] And the search for a correct technical implementation of algorithms has to be counted as the second main task of computer science.

As already mentioned in Section 1.4, communicative and social competence is the keystone for every computer scientist to her or his professional success; without the abilities (1) to consider the non-technical preconditions and implications of a technical problem at hand, and (2) to communicate the quality of a proposed solution to those affected by the consequences, the best structural and technical knowledge is not enough to make a good computer scientist. Taking »data structure« and »implementation« as the central subjects of computer science rather than »data processing« and »computer« helps us to better understand how the methodological core of the discipline may interact with those "soft skills": the connection to the theory of rational argumentation explains in a clear way how an implementation relates to a certain understanding of pictures (for example) used in a particular field of application. The answer to the question 'What kind of "translation skill" is to be used in order to understand the problem to be solved?' is: "Listen carefully to grasp the fields of concepts structuring the argumentations in their domain." And the answer to the question 'How should the resulting computer systems be explained as the expected solution ("translated back") to the users from that field (who are not specialists of computer science)?' has the form: "Explain your implementation as a rational argumentation: Introduce the structure of your implementation as a combination of concepts already agreed upon, and show that that structure necessarily fits the specified criteria."

It is a crucial intention of computer science in general to provide by its results others with tools to deal with *their* problems – for the example of image-related software: the physician, the industrial designer, the material scientist, the historian of arts, the physicist or the creator of cinematic special effects among others. Therefore most questions and argumentations of those areas of application reappear in the "micro cosmos" of computer science. With respect to computational visualistics – i.e., the science of images in computer science – this is particularly true for the diverse concepts developed in the general science of images.

2.2 Visualistics and the Many Sciences of Pictures / Images

When characterizing visualistics in the introduction of this chapter as "a new unified science of pictures" , we of course have no intention of denying that there have been – indeed for a long time already – numerous sciences of pictures occupied with the description and analysis of pictures and picture uses from various points of views and with diverse methodologies. Although quite common nowadays, the expression in singular "science of images" dates actually back only to the 1990's prepared by several calls articulating the need for such a new approach – with variant expressions: "imagic turn" [FELLMANN *1991, 26*], "pictorial turn" [MITCHELL *1992, 89*], "iconic turn" [BOEHM *1994, 13*] among the more well-known.

The scientific subject of general visualistics is given by any form of images and pictures: esthetic images of art and functional pictures of advertisement, graphics in mathematics and visualizations in medicine, Indian sand pictures and computer-generated 3D-environments, *trompe l'œil* paintings and airport pictograms, children's scribbles and masterly Paleolithic cave paintings, failed photos and excellent video-recordings. The characterization used to bind together these quite different phenomena

[3] This is, of course, a variant of the famous CHURCH/TURING thesis [KLEENE *1967, 232*]

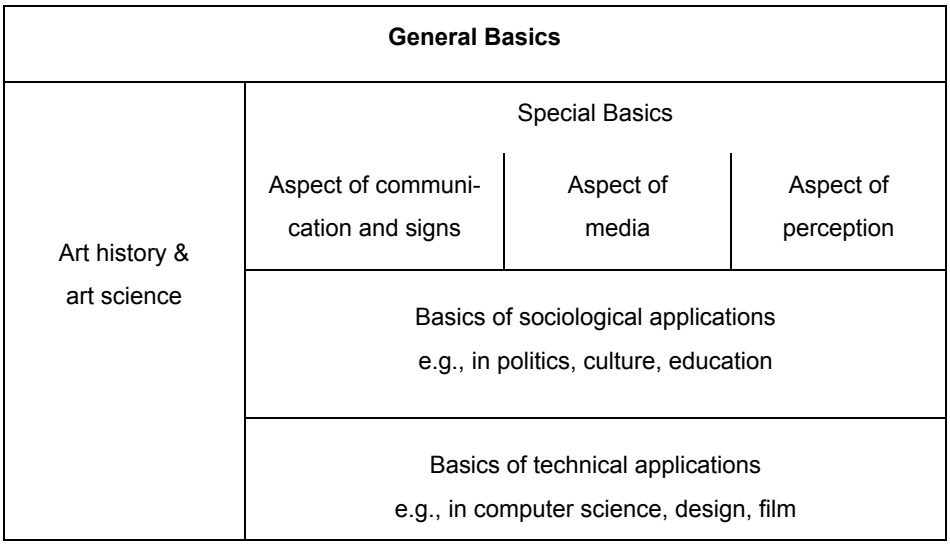


Figure 1: SACHS-HOMBACH's Organization of the "Scientific Parts" of Image Science

in the new approach of general visualistics is their classification as *perceptoid signs* [SACHS-HOMBACH *2001, 18ff*]: each of them is – or is at least intended as – a tool for communication (»signs«) that has to apply our abilities of visual perception in a specific manner (»perceptoid«) in order to function properly. More precisely, in using – i.e., adequately using – pictures we do not only perceive visually the sign in its physical appearance (which would be the same with reading written texts): we have also to invoke – at least to some degree – our abilities to visually perceive spatial objects and configurations that are closely related with what the picture is employed to symbolize. We shall come back in Chapter 3 to a more precise discussion of this definition of images and the consequences it bears.

SACHS-HOMBACH [2005] bases a collection of essays on general visualistics and its relation to the partial sciences of pictures and images on the following grouping of the participating disciplines (Fig. 1):[4]

- those concerned with the theoretical foundations (including (in alphabetic order) essentially cognitive science, communication science, mathematics, neuro sciences, philosophy, psychology, science of art, and semiotics);
- those orientated historically (archeology, ethnology, history, and museology);
- those in the context of social sciences (cultural science/visual culture, education science, media science, political science, sociology);
- those considering or enabling applications (advertisement, cartography, computer science, typography);
- and those producing various forms of pictures (art, design, film and TV, photography, digital media).[5]

As for the methodology of general visualistics, the interdisciplinary background opens a broad range of methods to be used while investigating perceptoid symbols. The philosophical roots contribute theoretical analyses. Science and history of art add more

[4] Taken from the „information for the contributors", personal communication; cf. [SACHS-HOMBACH *2002*]; cf also [SACHS-HOMBACH & REHKÄMPER *1999*].

[5] The list is by no means intended as being complete.

descriptive-hermeneutic approaches. Design and also computational visualistics complement with a constructive component.

While the efforts toward an integral science of images at Magdeburg evolved around the double center of philosophical picture theory and computational visualistics, complementary approaches have followed that take mainly history of arts and science of art as a starting point and combine them with considerations from cultural anthropology (cf., e.g., [BELTING *2001*]). As we are mainly interested in the relation to computational visualistics, the Magdeburg approach is more directly useful, and thus, bases the sketch of image science in Chapter 3.

2.3 Computational Visualistics and the Data Type »Image«

In computer science, too, considering images and pictures has originally evolved along several more or less independent questions, which lead to proper sub-disciplines: computer graphics is certainly the most "visible" among them. Only just recently, the effort has been increased to finally form a unique and partially autonomous branch of computer science dedicated to images and pictures in general, and named 'computational visualistics' in analogy to computational linguistics.

For a science of images within computer science, quite obviously the abstract data type »image« (or perhaps several such types) stands in the center of interest together with the corresponding data structure (s) and the potential relations of implementation. Keeping the distinctions of section 2.1 in mind, a reasonable methodological *ad hoc* organization of the field could be derived by distinguishing the examinations of computational visualistics along the following three paths: we may be interested (a) in a purely field-internal consideration that concentrates exclusively on the abstract data structure around the type »image«, the basic operations that determine the structure, and the algorithms that can be defined with those operations; or (b) in the relations of implementation that may lead from more elementary data structures to the structure with the type »image«, and that would allow us to technically implement the image-algorithms of particular value for us; or (c) in the relations of implementation that open up even more complex data structures on top of the one including the type »image«, e.g. in VR systems. The considerations in chapter 4 follow essentially the first path.

Each of the "traditional" image-related sub-disciplines of computer science considers those three methodological aspects to various degrees. The distinction establishing the disciplines follows a simpler semantic pattern resulting from the types of operations and algorithms around the data type »image«, which relate an instance of »image« with something that either is or is not of the same type. From this criterion the following three main fields result (cf. Fig. 2) – we only give a short overview at this point:

- *Algorithms from »image« to »image«*
  In the field called *image processing*, the focus of attention is formed by the operations that take (at least) one picture (and potentially several other parameters that are not images) and relate it to another picture. With these operations, we can define algorithms for improving the quality of images (e.g., contrast reinforcement), and procedures for extracting certain parts of an image (e.g., edge finding) or for stamping out pictorial patterns following a particular Gestalt criterion (e.g., blue screen technique). Compression algorithms for the efficient storing or transmitting of pictorial data also belong into this field.

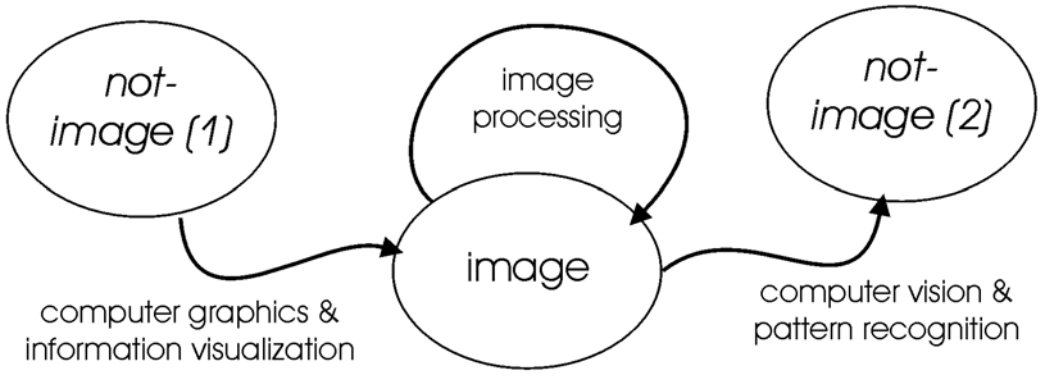


Figure 2: Sketch on the Operations with the Data Type »Image«

- *Algorithms from »image« to "not-image"*
  Two disciplines share the operations transforming images into non-pictorial data types. The field of *pattern recognition* is actually not restricted to pictures, but it has performed important precursory work for computational visualistics since the early 1950's in those areas that essentially classify information in given images: the identification of simple geometric Gestalts (e.g., "circular region"), the classification of letters (recognition of handwriting), the "seeing" of spatial objects in the images or even the association of stylistic attributes of the representation. That is, the images are to be associated with a non-pictorial data type forming a description. The neighboring field of *computer vision* is the part of AI (Artificial Intelligence) in which computer scientists try to teach – loosely speaking – computers the ability of visual perception. Therefore, a problem rather belongs to computer vision to the degree to which its goal is "semantic", i.e., the result approximates the human seeing of objects in a picture.
- *Algorithms from "not-image" to »image«*
  The investigation of possibilities gained by the operations that result in instances of the data type »image« but take as starting point instances of non-pictorial data types is performed in particular in *computer graphics* and *information visualization*. The former deals with images in the closer sense, i.e., those pictures showing spatial configurations of objects (in the colloquial meaning of 'object') in a more or less naturalistic representation like, e.g., in a computer game. The starting point of the picture-generating algorithms in computer graphics is usually a data type that allows us to describe the geometry in three dimensions and the lighting of the scene to be depicted together with the important optical properties of the surfaces considered. Information visualizers are interested in presenting pictorially any other data type, in particular those that consist of non-visual components in a "space" of states: in order to do so, a convention of visual presentation has firstly to be determined – e.g., a code of colors or certain icons. The well-known fractal images (e.g., of the MANDELBRODT set) form a borderline case of information visualization since an abstract mathematical property has been visualized.

The algorithms behind the arrows in Figure 2 may indeed consist of complicated combinations of all three possibilities mentioned above: For example, we may consider a procedure in computer graphics that is put in sequence after an algorithm of computer vision in order to solve a complex problem in image processing. Within this framework,

investigations have a focus on structural aspects or on engineering problems – mirroring the traditional differentiation between a more mathematically oriented theoretical informatics, and the engineering-oriented practical and applied computer science.

The interdisciplinary structure of computational visualistics influences its methodology, as well. The main focus is on the constructive side. But the clear understanding of the underlying data structures requires at least a profound overview of the methods of the other disciplines of visualistics. Indeed, the work of computational visualists may be considered as of the following three essential components: compiling partial specifications of a data structure the implementation of which is needed by a client. Augmenting the – probably incomplete – specification in a coherent manner. And finally, implementing the specification either in the abstract or the technical sense, or mostly both, so that the client can apply the data structure initially specified in an automatized manner. The second and third tasks, being field-internal and field-external considerations respectively, are what has traditionally been thought of as the central work of computer scientists that does in fact not change much for different fields of computer science. The first task holds the true domain-specific aspects. For computational visualistics, the argumentations of image theory provide the necessary clarifications.

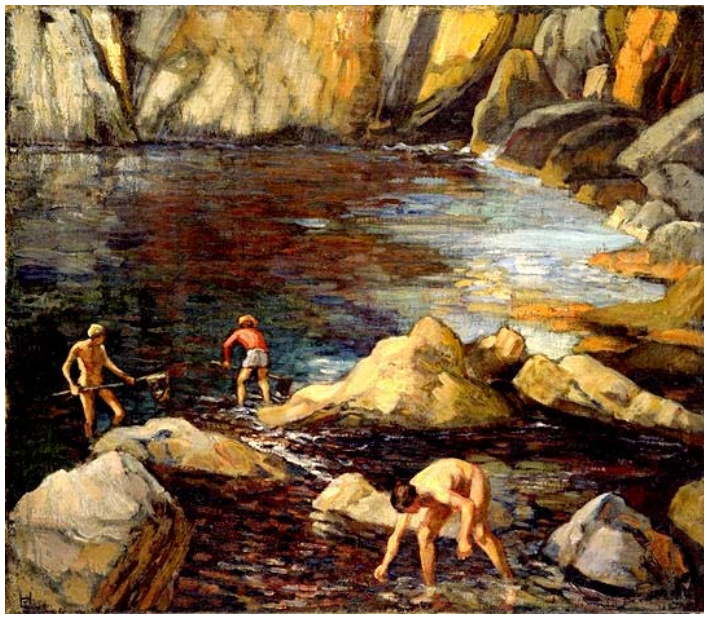
Figure 3: Where does the picture end?

Art Imitating Life Imitating Art Imitating Life. JOHN PUGH, deceptive mural, with framing brick walls

3 Preliminary Clarifications from Visualistics

Pictures seem to be a very easy and simultaneously a very complicated subject of investigation: on the one hand, nobody has serious problems in everyday life to distinguish pictures from other things, and to use them exactly as pictures. On the other hand, it remains notoriously unclear even in scientific contexts, where (or better: how) that border is to be drawn, or with what internal characterizations the manifold of context-dependent uses could be systematically explained in a satisfying manner. In the following, a condensed overview on crucial aspects of image theories in visualistics is given as an introduction to every computer scientist interested professionally in pictures. Correspondingly, there are few references to computers and data structures in this chapter.

We first (3.1) have a superficial look on a collection of borderline cases that may render us more sensitive for the reach of the class "picture", for its less typical subcategories, and for the erroneous properties we easily attribute to the concept »image« from our colloquial but too narrow understanding. Section 3.2 introduces and elaborates the conception of images as "perceptoid signs" that is central for modern visualistics [SACHS-HOMBACH *2002, 53ff*], hence also for the rest of this book. In this framework, investigations on the relations of images to the objects depicted (3.3), to the communicative functions of verbal language (3.4), and to the picture users (3.5) are presented.



(a) L. V. HOFMANN, *Fischende in Felsenbucht*, ca. 1910

(b) Photograph of the author

Figure 4: Examples of a Picture of Art (a) and of a Private Photograph (b)

3.1 Pictures on the Border: Overlooking a Wide Kingdom

Asked to name spontaneously a picture just coming to mind most people mention personal photos (holiday snapshots, passport portraits), pictures of art or advertisment (genre-paintings, billposters) and illustrations in books or papers (weather charts, scientific graphs). At the core of the concept are, we might conclude, flat smooth material objects a marked surface of which shows permanently the significant distribution of pigments. But beside those "normal" pictures (cf. Fig. 3, and again Fig. 2), there are less central cases: what about TV pictures, projected images from a slide, stained glases, mirror images? What about the optical image on the retina? In the following, a "gallery of the curious" of unusual or even questionable pictures may broaden our view.

Perhaps with the exception of the last case (and this exclusion indeed holds only on first view – think of an ophthalmologist), all the examples given above are things to be *seen* – they belong to optical phenomena that have to be visually perceived by (or at least perceptible for) somebody, and thus connect the physical dimension with the mental one. Although the expressions 'image' and 'picture' are also used for phenomena that are accessible by other modalities of sense (or even for verbal metaphors), which at least partially qualify for the definition of »image« as perceptoid sign given below as well (Sec. 3.2), we exclude in the following all cases that are not predominantly visual.

Usually, we understand a frame as being a necessary (external) part of a picture: a border marking which part of the total surface of a "picture vehicle" is to be considered "being in the picture". As in the examples of Fig. 4, this border may consist only of a discontinuity in pigmentation in a standard shape (predominantly a rectangle in European tradition), but there may also be an explicit frame – additional lines or special physical devices to emphasize this border of the pictorial space. The frame indeed marks one figure-ground distinction associated with pictures. A second one applies *in* the picture's space: the distinction between the image's foreground objects (e.g., a per-

Figure 5: Pictorial Tessellation

M.C. ESCHER: *Eight Faces*, 1922

Figure 6: Cubistic Specimen

J. GRIS, *Portrait of PICASSO*, 1912

son in Fig. 4b) and the background in front of which they are depicted/perceived (bits of meadow and grove in Fig. 4b). This is usually a much more fluid distinction, similar to ordinary visual perception: we can see a crouching human figure in front of a landscape in Figure 4a, but we may also separate the rock in the middle from the surrounding scene in our perception.

There are exceptions to both types of figure *vs.* ground for pictures: in pictures with a tessellation,[6] like the well-known works of M. C. ESCHER (Fig. 5), an observer's attribution of figure and ground in the image space changes more or less involuntarily, depending on where they focus their attention on momentarily. Pictures without a frame are quite common in the form of highly naturalistic representations intended to deceive the beholder's eye (cf. Fig. 3), traditionally named in French: *trompe l'œil* – 'deception of the eye'. Seemingly (at least on first view), these pictures lose their "pictoriality": there appears to be a real statue on the left side of the alcove, and a real girl sitting and reading at a table on its right side in Figure 3. Of course, here in the book, printed in small format and in gray values only, what is given is indeed the image of a picture – which is also true for most of the other example pictures shown here (even often with several further intermediate steps of representation). Like ordinary quotation of words and phrases, "pictorial quotation" obeys special rules as to which aspects of the picture quoted remain unchanged (e.g., proportions, intensity), and what others may be left apart (e.g., color, size; [STEINBRENNER *1999*]).

Like pictorial quotation, pictures of art often emphasize certain aspects of being a picture or using a picture. For example, pointillist pictures are often interpreted as guiding our focus of attention to the theory of coloring (and the schematic treatment thereof in earlier academic painting, among other factors); cubistic works of art draw our attention to the fact that spatial objects always integrate a multitude of perspectives not just one (cf. Fig. 6); nonfigurative art demonstrates in various ways that pictures are not only used to represent spatial scenes (with traditionally associated cultural significance).

[6] Tesselation: the geometric plain is fully covered with non-overlapping segments in an iterative manner.

Figure 7: Drawing of Maori Facial Tattoo for Chieftains

Figure 8: Australian Aborigine Tschurringa: A Map as Mythological Proof of Territorial Property

One such non-representational use appears in body paintings, in so-called primitive tribes as much as in "civilized" cosmetic face painting or (more or less temporary) tattoos (Fig. 7): these border cases of pictures clearly serve primarily as a means for self-portrayal of their bearer, i.e., their tendency to react in certain manners to particular conditions, or at least to be seen as such. Avatars in 3D-interactive virtual meeting places[7] and their changeable "skins" have indeed an analogous function. Although the self-expression by means of body paintings (or avatars) does not have to be always sincere, the distinction "true or false" is here as inadequate as it is for masks (which we also take as a border line case of »picture«). In all those cases, the picture "screen" is not flat, and the patterns of pigmentation used are often highly abstract or schematic, emphasizing bodily features or indicating certain gestures or mimic.

Combined with a more traditional kind of reference, a high degree of abstraction is also to be found in maps: geographic features and/or passages (or obstacles) for traveling are represented in a variety of highly stylized and culture-specific forms. In many early cases of map usage, the property of the map was ritually linked to the control of the corresponding territory (Fig. 8) – a habit still vivid in the maps and registers of land registry offices in our more urbanized societies. A similar picture, but with even more (and a different type of) abstraction involved, is given by illustrative sketches as in Figure 2: that "geography" is indeed completely unreal, the geography of a field of concepts, so to speak. We might also say that such a picture presents the passages one's argumentation may follow. In a way, to have that image is to control that knowledge, as well.

Does the pattern of a Scotch kilt qualify as a picture? As we have seen so many deviations from the naïve determination of the concept »picture« so far – flat or not, with or without frame, with or without a unique figure-ground-distinction, with or without referential links to real entities – there seems to be little sense in excluding such color patterns from being considered pictures. Even more so, if they are conceived of in their traditional function of indicating a family membership, which brings them functionally very close to the body painting examples mentioned above. In general, decorative elements and ornaments are often derived from a representational original (cf. Fig. 9). Some elements of a representational picture are isolated, graphically simplified, and then used repeatedly, e.g., as an ornamental border of another picture. Although their

[7] cf. for example http://www.atmospherians.com/at_avatars/avatars/listings.html

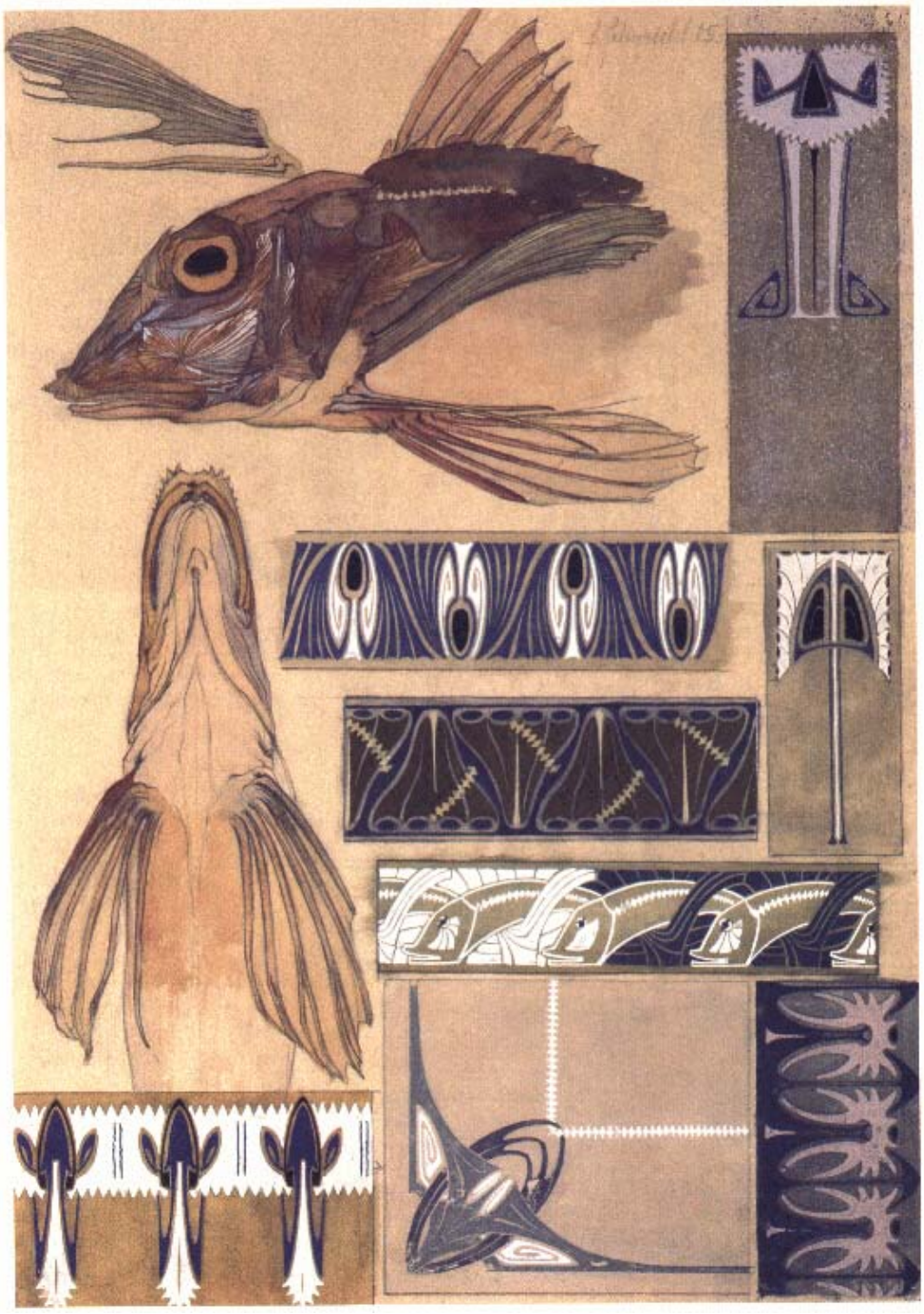


Figure 9: Example for Developing Decorative Elements from Representational Pictures

R.B. SCHURICHT, *Naturstudie*, 1905

origin may not be recognizable later, such decorative elements may throw a kind of dim meaning halo on other picture elements, enhancing a certain interpretation or coloring of the general impression.

This, by the way, is also a crucial ingredient of traditional Chinese poetry: the ideograms forming Chinese writing not only encode words; they are composed of graphical

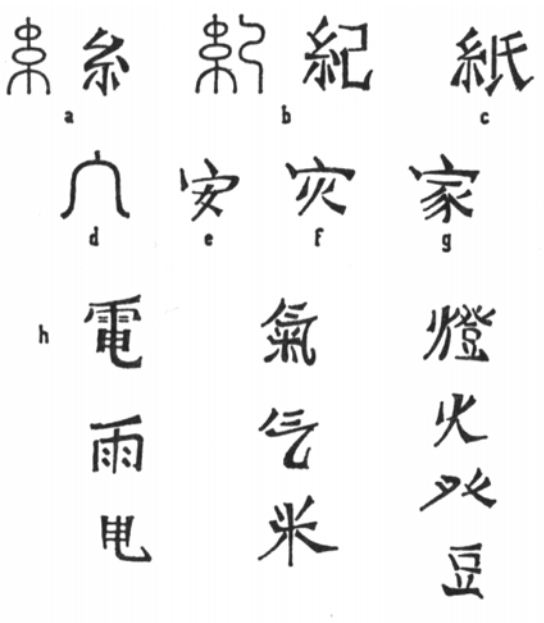


Figure 10: Some Chinese Characters

a) archaic and modern version of character “bundle of fibers, thread”; b) combination of the thread with the movement of a shuttle (archaic and modern version) meaning “order, sequence”; c) the thread combined with phonetic component (“paper”); d) archaic pictogram for “roof”; e) combination “women under roof” = “peace”; f) “fire under roof” = “accident, mishap”; g) “pig under roof” = “family”; h) three character word “light bulb” (left to right) with respective elements (below): “rain – flash”; “steam from a pot – rice”; “fire – rising – base”;

Figure 11: *Cloudy Mountain After Rain*, CHITFU YU, 1997

lowest black character = mountain; in upper right corner (gray) character for rain (cf. Fig. 10h)

elements derived by simplification from graphics with quite a direct representational association (cf. Fig. 10). For the connoisseur these connotations are still visible and form a halo of weak additional meanings modifying the literal meaning of a poem – an effect held in high esteem (and barely comprehensible for somebody used to writing based merely on phonetic letters). In Chinese calligraphy, the roles are exchanged: it is the literal meaning of the written words that adds an unusual component to the understanding of a primarily graphical-expressive painting (Fig. 11).

In all the examples presented so far, the beholder can repeatedly have looks at the picture over and over again. Indeed, most of those pictures only work as intended if the beholders really have several looks at them at different times. The pictures are of a *persistent* nature. Though in many tribes of Australian and American indigenous people, a frequent means of cultural expression are sand drawings. Such pictures are “drawn” by strewing colored sand in patterns on a relatively flat part of the floor, or by pushing lines and dots with a stick or the fingers in flat monochrome sand or mud (Fig. 12). They are usually produced in the course of a religious ceremony, which also requires the picture being destroyed at the end. As for the pictures produced in a *life* TV broadcast (without recording), these images are seemingly not persistent, too, and cannot be accessed after the event.

But then, when the same ritual is performed again, the members of the culture insist that the *same* sacral picture is brought into appearance. It is the material picture vehicle

Figure 12: Photography of an Australian Sand Drawing

Figure 13: Simultaneously Sculpture and Mask Screenshot from an Insect-like Avatar

that needs not be persistent, while the picture *per se* still continues to exist – so to speak – and may come forth materially again in a different context. We here touch the question of the identity of pictures that has been discussed quite controversially among picture theorists. It may suffice at this place to say that a picture is best being conceived of as an abstract entity. In some cases this entity is considered as being bound immediately to its material vehicle, and thus disappears together with the latter (e.g., we would think of "Las Meniñas" as irretrievably lost if the famous screen is destroyed, leaving behind mere copies). In other cases the same picture may be materialized (successively or simultaneously) more than once.

Sand pictures drawn with a stick or finger are not smooth – in particular the shadows thrown by their three dimensional structure are indeed necessary for perceiving them. Similarly, in some artistic styles, "texture" includes a three-dimensional distribution of pigments that contributes shadows as an essential ingredient to the pictures. Correspondingly, relieves and engravings (the plates, not the prints!) depend on being perceived visually under certain illumination conditions, although they also may be perceived haptically. As the former is the major path of access to them, we shall consider them as another peripheral case under the concept »picture«. Following that path even further, we wonder whether sculptures should be included under the concept – as the German expression 'Bildhauerei' (literally: 'image hewing') for sculpture's art clearly suggests. For our purposes it is certainly advisable to include them at least as marginal cases of pictures, since we do not want to block the possibility of studying cases of *virtual reality*, i.e., highly interactive computer graphics, as much under the perspective of the two-dimensional projection as under the viewpoint of three-dimensional modeling – the former binding the investigation more to the center cases of pictures (in particular to *trompe l'œil)* while the latter connects it to sculpturing and architecture. An avatar (Fig. 13) appears at each moment as a picture, but in order to be able to generate those instantaneous pictures, it has to be modeled like a sculpture first.

Let us finish our small "gallery of the pictorial curious" with considering a really special find: Do we have to classify the marks made by chimpanzees in some experiments as images? The most famous of such events is described by [GARDNER & GARDNER 1980]: The captive chimpanzee called 'Moja' was trained to communicate with American Sign Language signs. On an occasion, the animal made some traces with chalk on paper (Fig. 14). A research assistant who had observed this behavior, signed immediately afterwards to Moja 'what that?' provoking the gestured reply interpreted by the as-

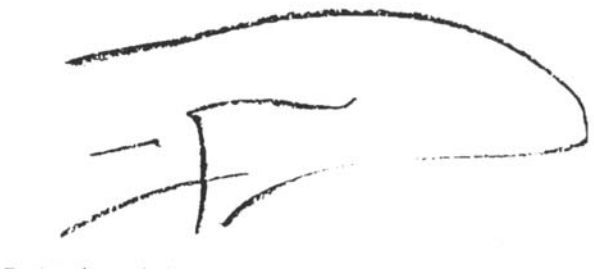


Figure 14: Moja's „Bird Picture"

sistant as 'bird'. Indeed, the question rather must be for *whom* is the paper with Moja's marks a picture. For the animal? For the assistant? For both? For us (from the distance)?

It is not a particular property of a picture candidate that decides the issue of its being a picture or not, but a complicated relation involving the object, the users and their context of action, and also their level of reflection on what they are doing there. In order to theoretically grasp the multiple harmonics of this orchestra, LOPES [1996] identifies two different (philosophical) tracks of discussion: influenced by linguistics and semiotics, pictures are viewed by some researchers, as a particular kind of sign, forming semiotic systems like a language.[8] Rooted in psychological theories of perception, pictures are conceived by others as a special phenomenon of (visual) perception, LOPES suggests.[9] Unfortunately, the "semioticists" often bind their investigations too closely to another particular type of signs – verbal language – and thereby ignore the special relevance of perception for pictorial signs, while the "perceptualists" have a drift to not bother with the communicative context every picture is used in; their weaker representatives sometimes even confuse perception in general with perception of pictures and put pictures in a categorical opposition to language and communication.

3.2 A Synthetic Proposal: Images as "Perceptoid" Signs

Backgrounded by LOPES's analysis of the historical situation of theory formation, the main methodological characteristics of the general science of images (visualistics) presently in *statu nascendi* can be summarized as a proposal to systematically combine the two lines of tradition by conceiving pictures as *perceptoid signs*.[10] Traditionally, a concept can be determined by giving a superimposed concept, and then adding the specific difference to the other subclasses. Applied to "perceptoid signs", we distinguish generic characterizations that pictures have in common with all signs from the specific difference »perceptoid«, which allows us to distinguish pictorial signs from other kinds of signs, e.g., verbal signs. There are a couple of important consequences of that conception, which guide us in the following in order to gain an understanding of what precisely is meant by "perceptoid signs".

3.2.1 »Sign« as *Genus Proximum* for Pictures

First: choosing »sign« as the superimposed concept clearly connects this position to the semiotic roots of picture theories. As a result, everything conceded about signs (in

[8] NELSON GOODMAN is usually conceived to be at present the most prominent "father representative" of the semiotic track; cf. [GOODMAN *1976*].

[9] ERNST GOMBRICH counts currently as the most influential and relevant "picture theorist" who follows mostly the perceptual track; cf. [GOMBRICH *1960*].

[10] It is mainly KLAUS SACHS-HOMBACH who has worked out this synthesis; cf [SACHS-HOMBACH 2001], [SACHS-HOMBACH *2002*]. He uses the German expression 'wahrnehmungsnahes Zeichen' – which is approximately 'sign close to perception'.

general) must be applicable to pictures, as well. In particular, they are embedded in a specific context of action – the sign act – that involves two participants: one role may be called the "sender", the other is the "receiver".[11] In a sign act, the sender *gives something to understand* to the receiver by means of the sign, that is, she acts in this specific way in order to direct the attention of the receiver to something. This "something" must – by logical grounds (cf. [ROS *1990, Vol. III, 125ff*] and [ROS *2005, 556f.*]) – be primarily an attitude of the sender: an expression of his/her/its readiness to be involved in certain reactions (cf., e.g., a bodily expression equivalent to "I am angry"). This primary attitude may or may not allow us to differentiate the understanding into the sender's intentionality toward a state of affair (potentially fictitious: a warning cry: "Tiger!!!"; or an assertion "In 1631 Magdeburg was completely destroyed") or toward an object (possibly not present: "the author of the *Philosophical Investigations*").[12] As a consequence, we never investigate single signs but always systems according to the multitude of "things" that the receiver is to be made aware of by means of the sign acts. Pictures are conceived of as a separate sign system – it may be decomposable in distinct subsystems: naturalistic representations, technical graphs, icons, etc.

Second: since "being a sign" is determined by the use in a particular type of activity it is obviously not a combination of attributes of a physical object that allows us to categorize something as a sign. We indeed produce sometimes artifacts with the sole purpose to be used as signs (e.g., name tags for the participants of a conference, street signs). But more or less any physical object may under certain circumstances be interpreted as a sign, as well. It is always the sign act and all its current (or potential, i.e., anticipated) participants that have to be considered if we understand something as a sign.

Third, if using physical objects as pictures is a communicative act, the various semiotic aspects are applicable, like the distinction between what is referred to, what is represented, and what is intended with the sign act (or, alternatively in BÜHLER's terms: 'representation', 'expression', and 'appeal' of a sign; cf. Fig, 15, and [BÜHLER *1933, 28*]).

In this context, SACHS-HOMBACH follows the tradition by suggesting to distinguish picture vehicle, picture content, and picture referent [SACHS-HOMBACH *2003*]: with the expression 'picture vehicle' we restrict our attention to those aspects of a picture that it has as a mere physical object, like a cathode ray screen, with the usual properties of physical objects including the visual ones of shape and color. As such, it may be employed in a sign act, but it also may be used in many other types of activities not related with communication. If we speak of "the referent of a picture" we mean the (factual or fictitious) scenes, events, objects, etc. that the picture is taken to represent. Finally, by considering the "picture content" we focus on those properties of the picture vehicle that are relevant for understanding its significance in the sign act.

[11] Alternatively, one sign user may simultaneously take both roles; e.g., somebody wandering through an art gallery contemplating the pictures ("showing them to herself"), or somebody performing soliloquy ("speaking to himself").

[12] Note that the first form is quite common among animals; the second form is restricted to higher vertebrates, the third form is private to human beings (or language users in the close sense of language). cf. [CLARKE *FC, Sect. 5*]. It is also important to note that, concerning the more complicated forms, it is still the sender's *attitude* toward a state of affairs/object that is denoted or referred to in a complex sign act: what is denoted or referred to and what is represented are usually *not* identical in a sign act (cf. [ROS *1989/90, Vol. III, 129ff*]).

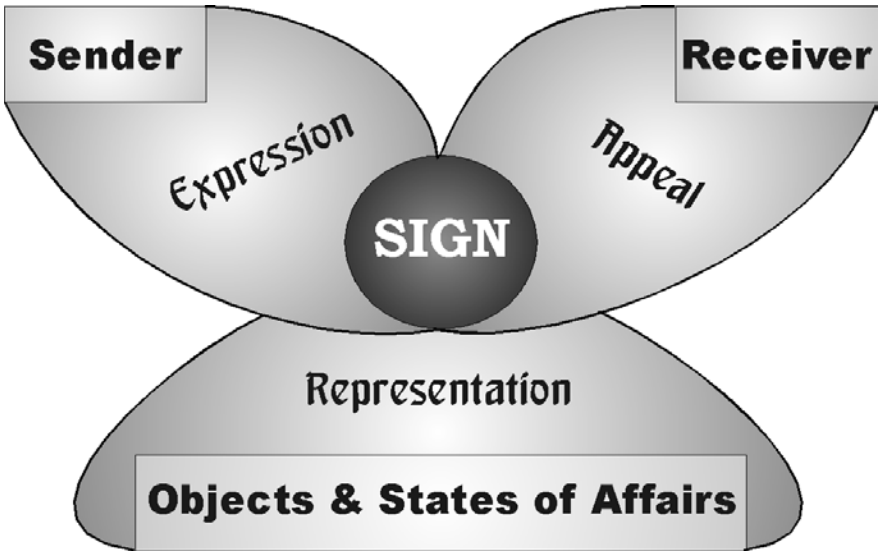


Figure 15: Standard Situation of Sign (S) Use (the *Organon Model* [BÜHLER *1933, 28*])

Fourth: denoting objects or events is not the only communicative purpose pictures are used for. By showing a picture to someone we might want to express our feelings or ask that person to do something (cf. 'expression' and 'appeal' in the organon model, Fig. 15). It is coherent with this understanding that the development of communications for infants is described as a sequence of repetitions and variations: based on innate emotional expressive acts, e.g., smiles, exchanged with (not just send to!) the reference person, variations are developed on both sides, and established or ignored by the mutual feedback reactions, so that more and more complicated patterns are formed that can be used to communicate more than just expressing attitudes [DORNES *1993, 152ff*]. Communication, then, may be seen as a kind of dance. Each of the diverse types of moves in a language game determines a corresponding interpretation schema for the concrete utterances, which cannot be associated with the mere sign vehicle used but is determined by the position in the overall choreography. For verbal language the sequence of pragmatically possible and plausible communicative moves in various language games is described by the theory of *speech acts* [AUSTIN *1962*, SEARLE *1969*]. Some authors have extended this approach to "picture acts" (including mixed communication, as well; [ANDRÉ *2000, Sect. 2.3.1*]).

Fifth: using pictures requires mastering a variety of semiotic rules: pragmatic rules that describe the typical functions and use conditions of pictures within the context of the other types of activities the communicating partners are (or may be) involved in; semantic rules that express the relation between the picture vehicle and its broader meaning as far as this relation can be construed without explicitly dealing with pragmatics; and – as the most abstracted level of investigation – syntactic rules that try to identify and formalize the range of attributes the picture vehicles must have in order to be usable as a particular sign of a system. The primary effect of the digitization of pictures necessary to deal with them in computers as mentioned above, is a syntactic effect, and we shall have to deal with that in more detail in the following chapter. But the two other levels have crucial influence on the way images are dealt with in computer science, too. Therefore the next sections of this chapter summarize some general observations on the particular relation (resemblance) between pictures and their prime referents (spatial objects), and on the relation between pictorial sign acts and the other acts of the picture

users in that context, in particular propositional communicative acts. We shall see in Chapter 4 that primarily semantic and syntactic features have been investigated in computational visualistics for most of its (pre)history; taking explicitly a pragmatic perspective is essentially a rather new development in the field.

Let us now consider the properties that distinguish pictorial signs from other subcategories of »sign«.

3.2.2 »Perceptoid« as *Differentia Specifica* for Pictorial Signs

What distinguishes pictures from signs like words and sentences is the special role perceptual competences play for constituting the picture's content, i.e., for interpreting the sign.

Sixth: *resemblance theory* provides one means of construing the characteristic role of perceptual mechanisms [REHKÄMPER *2002*]. According to that approach, a sign is a picture if the perception of essential properties – that constitute the pictorial content – is identical to the perception of the corresponding properties of some other object under a certain perspective. Thus, we may use an object *S* as a pictorial sign for something else motivated by the observation that *S* looks similar to that other thing; however, similarity is a secondary condition for being a picture working only within the semiotic context of use. GOODMAN's critique of resemblance as necessary condition of being a picture is indeed directed only against taking resemblance as a condition constituting the semiotic context of picture use instead of restricting the general scenario of sign acts in a particular manner [FILES *1996*].

Seventh: talking about likeness between perceptions instead of resemblance between objects brings the psychological characteristics of (visual) perception into account, and with them the principles of object constitution underlying them. For example, psychophysical restrictions of receptivity, laws of Gestalt formation, conditions of color invariance, and factors of interpretative schemata for 3D-perception must be considered if we want to understand what objects appear as similar for somebody, and in what respect. As object perception may be distinct for different social groups (different by culture or by age), so may be the motivation to use and understand an object as a pictorial sign.

There is no general restriction to the visual sense: perceptoid signs may be conceived accordingly in any modality. This accounts for the use of the expression 'picture' for non-visual perceptoid signs. Whether we can use such "pictures" to focus our attention to things in a way similar of using (visual) pictures depends in these cases on whether there are methods of object constitution associated with the sense modalities involved: while it is relatively easy for us to employ "sound images", we usually would need some training in order to use "odor images" for more than evoking very generally a situational context.

Eighth: the shift to psychology also opens an interpretation of pictures with fictitious objects or contradictory scenes by means of resemblance theory. The objects of perception are "intentional objects" [HUSSERL *1980*], i.e., objects as something in the mind, something one's attention is directed to, something "in one's intention"; not something existing independently of any such intention and anybody having the intention. What we perceive in the case of an optical illusion, for example, cannot be – by definition of 'optical illusion' – an "objective object." Correspondingly, there need not be a likeness to any real object or scene for a corresponding picture, only one to intentional objects or imagined scenes.

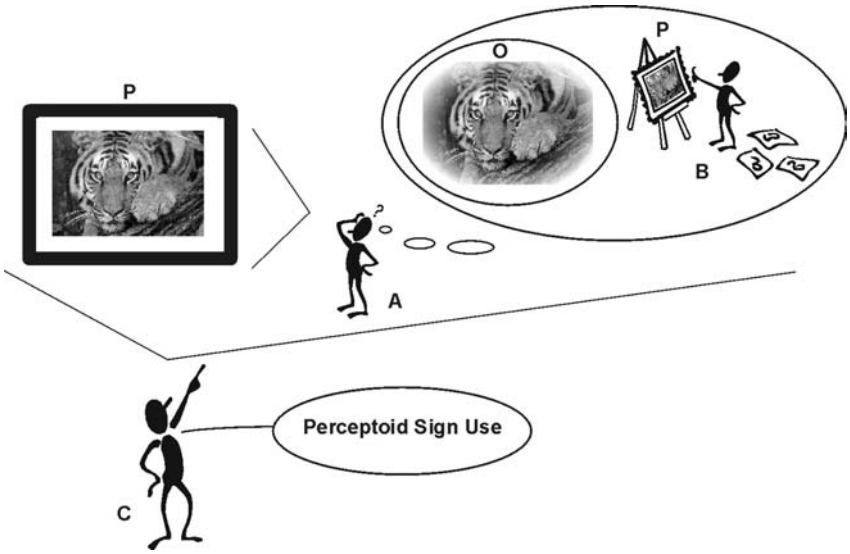


Figure 16: Perceptoid – the Special Connection to Perception for Pictures

Ninth: resemblance as a criterion to characterize pictorial signs presupposes that certain properties are excluded that do not contribute to the content of the sign and are irrelevant to its interpretation. Regarding something as a picture makes it obviously irrelevant how heavy that object is or what its back looks like. In the case of linguistic expressions, we consider some respects as irrelevant, too, but different ones, e.g., the color of the font. Resemblance comes as a vague criterion; but the fact that it is determined only in certain respects allows us also to accommodate it to quite different pictorial phenomena. In some cases – mainly naturalistic pictures like photographs – it seems that we immediately and involuntarily regard most respects as relevant that are also relevant in perceiving objects visually. In others, like line drawings, we leave aside many of those respects: the picture is taken to resemble some object only relative to the remaining respects. In the extreme case, a diagram, for example, does not show us anything about how any physical object *looks* like.

Tenth: it is possible to establish different respects of similarity as dominant because the picture vehicle does not in itself determine which properties are relevant for the depiction. We may develop[13] different pictorial schemata with respect to some particular communicative functions the pictures are supposed to perform. It is then true to say that all pictures resemble their objects in one way or the other, but this relies completely on the pictorial schema determining in each case the relevant respects. Even more: actually perceiving resemblance depends then completely on the use of the sign vehicle as a sign ruled by a particular pictorial schema. Resemblance can only be established as embedded in the sign act (cf. Fig. 16).

Eleventh: in consequence, the distinction between pictures and their content is less clearly marked than is the case for language. The visual impression of a *trompe l'œil* is – as a crucial feature of this type of image – more or less identical to the impression of the real object depicted in that picture. This closeness to the content gives us the impression of an access that is intuitive: we have to learn with an effort to master words, but to understand pictures seems to be a congenital facility for humans. Compared to verbal language perceptoid signs are less conventional, though the range of conventionality covered may still be rather broad.

[13] – within one set of principles of object constitution, see sixth item.

Figure 17: A Shadow in Hiroshima — August 6, 1945, 8:15 a.m.

Burn pattern on the steps of the Sumitomo bank building – the only indication a human life has left in the moment the human race demonstrated its power of ultimate self destruction

Twelfth: This also helps to understand the strange double fact: that we can on the one hand communicate with pictures in a way much more precise and more immediate than we are able to do with verbal language; but that we on the other hand usually need more contextual information to disambiguate what is actually meant by a picture from a lot of possible interpretations. In the words of SACHS-HOMBACH [SACHS-HOMBACH & SCHIRRA *1999, 35*], their high degree of *semantic abundance* comes along with a significant lack in *semantic precision*. Therefore, pictures need to be used in a context of action that determines in the concrete event their meaning, e.g., most explicitly by employing a caption.

3.2.3 A Note on "Natural Images", "Indices", and "Icons"

We have mentioned in the beginning of this chapter the images in mirrors. The phenomenon as such is obviously independent from any context of sign use. So the question arises whether – or in what respect – we can classify mirror images as perceptoid signs. Or put inversely: is the definition of pictures as "perceptoid signs" too narrow to include mirror images – and hence probably too narrow in general? How do we have to interprete "natural signs", as mirror images are sometimes called together with object shadows (cf. Fig. 17), the red spots of measles, and the foot prints on a sandy beach? In fact, in a society of blind nobody would have the idea of associating the discourse about, e.g., the surface of a quite lake with the discussion on perceptoid signs: considered as a mere object without anybody (even potentially) perceiving it in the right modality of sense (i.e., visually), there is no reason to link a mirror with pictures, at all.

C. S. PEIRCE has introduced the semiotic distinction between "index", "icon", and "symbol" that comes in handy for this discussion [PEIRCE *1931ff, 2.274*]. An index is an entity that may be used as a sign for something – the referent – due to its direct physical relation to that referent. Thus, we may use smoke as an indexical sign for fire – we keep

our focus of attention on that (assumed) fire by means of "showing to ourselves" the smoke we perceive alone. Analogously, a photograph may be conceived of as an index. Due to the causal relations – mediated by the light energy and the chemical reactions of the photosensitive emulsion – between a spatial scene and a photo thereof, a person can use the photo in a sign act to move the attention of somebody else to that spatial scene (or rather his/her attitude toward that scene).

Indexical signs cannot be used to refer to fictitious scenes – taking something as an indexical sign implies the reality of the referent. Nevertheless it is possible to lie with a photo [HGBRD *2000*]: if the sender is aware of the photo not to be an index (e.g., being a photomontage) but leaves the recipient in believing it to be used as an indexical sign this sign act fulfills all criteria of a lie. That a photo can be used as an indexical sign *and* as a non-indexical sign for referring to the very same scene leads us to PEIRCE's second class: icons are objects that may be used as a sign for something motivated by the fact that they bear resemblance with the intended referent. In the case of a photo, such a visual resemblance is usually assumed even in the case of massive alterations of the original index: then, a fictitious scene is assumed to look like that, and the photo may be used in a sign act to denote that fictitious scene. Films with naturalistically rendered computer graphics that place believably behaving dinosaurs together with real actors in the background of an exotic forest, which may or may not be a real landscape, give a perfect example of such an iconic sign of a visual fiction.

PEIRCE's third class, symbols, are characterized by no such immediate relation between sign vehicle and sign object or sign content: it is the semiotic activity of the sign users in general that is responsible for that connection. Hence, symbols are called *arbitrary*. For indexical signs and iconic signs it is possible to understand them without learning – by spontaneously activating knowledge about causal relations or resemblance within the contextual semiotic activity: "this guy tries to tell me something (anything!) with that thing, which looks similar to / is causally linked to …". No such spontaneous *semiosis* may take place for symbols without a prior introduction. The meaning of words, "human life" for example, must be taught; the significance of a date, e.g., "August 6, 1945", must be explicitly communicated (as part of an already established complex cultural frame) before they can be used as symbols.

In the light of these distinctions, we can interpret mirror images as icons and as iconic indexical signs: the situation for the interpretation as an icon is given when we get a fright because taking erroneously our own mirror image – in the periphery of our field of sight, or when it's a bit dark – for another person appearing there unexpectedly (PEIRCE's "genuine icon" [PEIRCE *1931ff, 3.362*]). If we realize a moment later what has happened, the mirror image changes its character immediately and becomes for us an iconic indexical sign: we focus our own attention to our own visual appearance by means of something looking similar and being causally linked directly to that appearance.

Quite obviously, iconic signs and perceptoid signs are closely related, to say the least. Note, however, that the explanations of "perceptoid" make an explicit reference to the psychological background of resemblance as something derived in perception according to the principles of object constitution. Speaking of icons does not necessary imply such a complication: if it is possible or favorable to define similarity between objects *per se*, iconic signs become possible that need not "feel" ("look" etc.) similar to the referent scene, as long as the perception-independent relation of similarity holds as well (and is used as a motivation for preferring that particular vehicle in that sign act). It is however

quite dubious that such a concept of resemblance apart from psychology should not be considered as merely derived and usable only within very limited conditions.

In any case, the use of iconic and indexical signs depends on the sign users' awareness of similarity or causal relation, which can be stabilized inter-individually *only* by means of symbolic communication. Without anchoring the language-mediated context of Figure 17, its iconic use remains ambiguous ("Is this meant as a human form or not?"), its indexical reference unclear ("where is this?" "who/what made that shadow-like spot?"). In order to provide a better understanding of how iconic (and indexical) sign uses depend on symbolically mediated frames of interpretations, the next section gives a coarse sketch of the complex inner structures of resemblance relations.

3.3 Image and Object

According to BÜHLER's organon model (cf. again Fig. 15, p. 24), *representation* is one of the three fundamental aspects of signs, characterizing in our case the relation between image and object (or state of affair) represented.[14] Although the obvious relation between image and object depicted is resemblance – seemingly a relation not too complicated to understand –, there has been considerable debate about its actual nature (and the consequences thereof for the concept »image«).

> *Almost everybody knows that the difference between a picture of Paris and a name or a description of it consists in the fact that the picture is more similar to the city. This of course is nonsense.*
>
> [GOODMAN & ELGIN *1988, VIII.1*]

The complication of similarity is most evident for pictures of fictitious objects. As a fast way out of this problem, the similarity between all the pictures of the same fictitious entities has been brought forward, though that is obviously not a real solution – at the least: one picture has to be the first. There are (for example in dictionaries) also pictures that are not meant to represent individual objects, but to demonstrate classes or concepts. Think also of the pictures on the doors of restrooms – they most certainly are neither (at least not in a straightforward way) similar to a class nor a concept. The resemblance to a typical member of the class depicted (or falling under the concept) has been considered to explain this case, but this, of course, extends any simple theory of resemblance in quite complicated manners.

So, what is resemblance? First of all, this (class of) relation(s) is not bound to pictures as one of the arguments; similarity may be stated for any two objects (states of affairs, etc.).

[14] Since "states of affairs" may be considered as a certain kind of object, too, only the expression "objects" is used in this section; states of affairs are explicitly dealt with in Section 3.4.

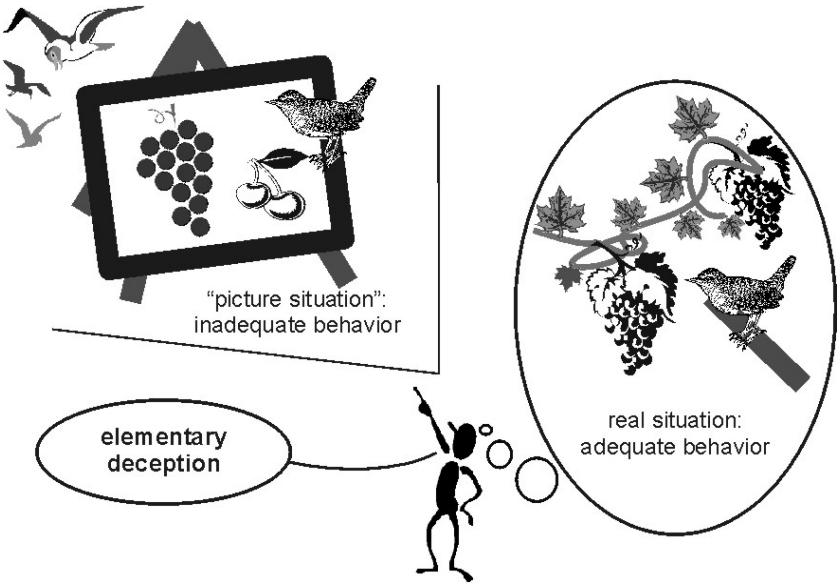


Figure 18: Ascription of an Elementary Deception: the Observer Assumes a Relation of the Observed Behavior to Another Situation

3.3.1 The Naïve Approach to Resemblance

> *We have a cat on the mat, and we have a picture of it. There is a part of the picture that corresponds to the cat; it is composed of parts that correspond in turn to parts of the cat: to the paws, the tail, the whiskers, ... This part of the picture is mostly black, except for one white bit that corresponds to a bit of the cat's throat; therefore the picture represents the cat as being black except for a white patch on the throat. The part of the picture that corresponds to the cat touches on another part of the picture that corresponds to the mat; therefore the picture represents the cat as being on the mat.*
>
> [LEWIS *1986, 166*]

The quote above examplifies (particularly with the final sentence) a common non-semiotic conception of pictures assuming a representation relation that is (*i*) independent of any use in a sign act, and (*ii*) immediately derived from a similarity relation. It demonstrates the naïve understanding of similarity as an objective relation between objects existing – without perspective – *out there*. In the naïve approach, the world consists of a set of individual objects – cats, mats, trees, chairs, teapots or bones – an observation quite in accord with everyday experience. These objects have specific distributions of attributes, which are used, assumedly, to classify or identify the objects by means of comparing them with an inner standard. Resemblance is conceived of, in this view, as a derived relation, a weak form of identity that may be stated to some degree if not all but only a certain amount of the attributes of two objects match. Therefore, likeness is one of the potential sources of deception (as in the case of the mirror image taken eventually for another person).

With a more sophisticated understanding, it is evident that objects are conceivable only as something given to an individual [UEXKÜLL *1909*]. That individual must then show a certain behavior, or act in a specific way, as the anecdote about ZEUXIS [PLINIUS *1977, 65*], a famous artist in ancient Greece, indicates: in the presence of a picture of fruit – perhaps ZEUXIS had exhibited them for sale by hanging the pictures in a tree –

birds behaved in a most peculiar manner. They mistook, it is told, this picture with what is depicted: they tried to eat the fruits "deceived by the similarity of the painting". But what is actually the criterion for such a statement? What clue do we have for speaking legitimately of "the bird's deception"? Well, the answer may be approximately like this: the birds flew hither and tried to peck up the feigned fruit – they behaved *as if* such fruit was *really present*. This behavior, or more precisely: *their* behavior that *we* recognize as not adequate to the situation while simultaneously imagining a situation in which it would be adequate for them is the condition that allows us to ascribe a deception to the animals (Fig. 18).

It is advisable to explain the concept "object" not as the concept of an isolated entity, but as a component of the concepts for certain dispositions to behave or act [PLESSNER *1928*]: an act-theoretic analysis of the concept of resemblance in the general framework of object constitution is therefore unavoidable.

3.3.2 The Act-Theoretic Basis of the Concept »Resemblance«

Studying systematically the relations between animal behavior and situational conditions is the subject of the biological subdiscipline of ethology. Ethologists can profit by experiments with situations mistaken by animals: they can, for example, use dummies with varying attributes in order to study relevant dependencies between corresponding aspects in an animal's environment and the actual behavior.

Mimicry, i.e., "imitating" something (in appearance and behavior) that their enemies ignore or avoid, is a "strategy" of numerous species to "protect" the corresponding individuals. Certain butterflies provide prominent case studies: if, e.g., an individual of the species *Smerinthus ocellata* feels being attacked it lifts its unspectacular brownish (camouflage) upper pair of wings and, thus, presents framed in a brilliant red its striking hind wings with their distinct black and light blue eye pattern (Fig. 19). The similarity of the lepidopteron's appearance with the head of a fox (or perhaps an eagle-owl) is hardly doubtable at least for the human beholder, and it is assumed that the main enemies of the insect (birds, rats, etc.) are sufficiently often impressed, too.

Figure 19: *Smerinthus ocellata* – a Case of Mimicry

Conceptually, behavior based on simple stimulus-response schemata allows us already to ascribe deceptions to a corresponding creature, though deceptions of a very reduced kind. Such instinctive behavior, called a "reflex" , includes the ability to distinguish the present situation of stimuli: those with a corresponding stimulus, and the others. The associated behavior is (usually) observed only in the former. The classification ability of reflexes is summarized by the formula "stimuli of the same kind lead to responses of the same kind[15]". The behavior of predators when facing successful mimicry is based on this merely *schematic ability of classification*. It is therefore wrong to assume that creatures thus endowed are able to deal with objects as *individuals* in the sense of our concept of material spatio-temporal objects – or even to be able to perceive two such objects as being similar. In the field of concepts of »reflex creatures«, there is no way of "projecting back" to the reflex arc whether the reaction performed was not

[15] – presuming the same inner conditions hold, i.e., apart form states of fatigue, illness, etc.

appropriate, i.e., whether the stimulus was not *really* "of the same kind" (cf. also [PLESSNER *1928, Sect. 6.2*]).

The behavior of ZEUXIS' birds is, however, not at all explainable by means of simple reflexes alone. There are at least two interacting reflex-like elements, since the birds do not just try to peck up the painted fruits; they have to come close, first: into pecking distance to the apparent fruit. It is part of the repertoire of such creatures to react on food already from a distance (much) greater than the one necessary for the "food ingestion" reflex proper to successfully "fire".

In this context, G. H. MEAD proposed to distinguish "distance experiences" from "contact stimuli". While reflexes in the proper meaning depend on contact senses (and lead to "contact reactions" accordingly), distance senses are understood as the sensory part of a reflex-like functional entity that is additionally related *in a systematic way* (i.e., by its concept) to other reflexes. Only in the later case, MEAD explains, it is legitimate to speak of 'perception' [MEAD *1932*]. As observers of creatures with distance senses, our attention is focused on activities that are (*i*) activated by stimuli of the distance senses, but that we (*ii*) understand as having the *purpose* of establishing (or avoiding) certain contact reactions (of the associated reflexes): we consider those movements in a broader temporal context than would be possible with the concept of reflexes alone. In the example of ZEUXIS' birds, the pecking (contact) reflex can be assumed as systematically associated to an "approach food" (distance) reflex – although the same stimulating object for us, there are involved two quite different and unrelated stimuli on the level of mere reflexes.

However, the punch line of the systematic relation is not only the approaching (or avoiding) reaction from the distance. That could be explained as a simple reflex (forming "a reflex chain"). For MEAD, the activation of the distance sense also provokes an *anticipation* of the contact reflex linked. Or more precisely: we must conceive the contact reflex as being *potentiated* before its stimulus is in fact present, so that the reflex can be activated quicker and more easily (weaker stimulus) – sometimes even without the actual stimulus (displacement activity).

Furthermore, with each distance stimulus, many potential contact reactions can be anticipated: "toward the water… – in order to drink, – in order to bath, – in order to cool down, – in order to prey," etc. Usually, a motivational structuring coordinates the diverse contact reflexes associated: e.g., the potential contact behaviors may inhibit each other, so that only one actually gets the better when the contact situation is reached. MEAD calls "resistance" the influence that the potentiated but finally not selected options have on the contact reaction actually performed. This resistance is in his understanding the origin of the primary object constitution [MEAD *1968, 413*].

3.3.3 Perception, Deception, and Primary Object Constitution

The systematic conceptual correlation of distance stimulus and contact stimuli prepares a first concept of an object that can be ascribed by an observer to such a creature. We can say then that for such a creature there exists – in contrast to mere stimuli – something like an object. That is, we can interpret the creature's behavior as being directed toward that object – though 'object' must be understood here in a rather rudimentary manner: this is but a precursor of our concept of objects as appearing in everyday language. The concept of such "pre-objects" (as they shall be called here for short) depends on the compound of anticipations "to one theme"; since a connection of several options of behavior is considered that remains invariant ("objective") compared

to the changing stimuli that appear from different distances or perspectives (cf. also [PLESSNER *1928, Sect. 6.3*]) .

The field of concepts coarsely sketched so far – let us call it the field of »pre-object creatures« – has been introduced by means of the field of simple »reflex creatures«. Instead of sensors, which react on stimuli in a certain bandwidth, we speak in this context of *detectors* for specific pre-objects: a detector integrates the corresponding distance sensors with the associated contact sensors. Similar to the mutual dependency of the concepts »sensor« and »stimulus« in the field of »reflex creatures«, the concepts »detector« and »pre-object« cannot be determined independently of each other: it is always a detector *for* a pre-object, or a pre-object *relative to* a detector. We may therefore say that *C perceives O* if a creature C's detector for pre-object O is activated.

Because the concept of perception is introduced in this primary object constitution as the fusion of reflexes, the corresponding reactions are part of »perception«, as well. A pre-object perceivable in this sense is something on which the creature can react in different ways – this is the basis of *properties* associated to the pre-object. Nevertheless, the pre-object is still "schematic": detectors always detect membership of a set, so to speak, not individualized objects with a single coherent spatio-temporal development. In certain cases, we may find a "detector for an individual", e.g., for parent binding; but then, this is merely a detector for a set that has just one member by chance, not by principle.

Pre-object creatures can be conceived of as having *intentional states*: the pre-objects to which these states are directed are intentional objects – objects *for* the creatures with corresponding active detectors. These intentional objects may even be "fictitious" : if, e.g., a food-detector becomes activated by distance stimulus – the creature perceives food from far – but there is not any stimulus for activating the corresponding contact reflexes for food when approached. Thus, the case of ZEUXIS' birds is explained quite well, although this determination of "fictitious" remains completely on the observer's side: The pre-objects are fictitious intentional objects *for us* – there is something looking similar to something else from the distance. *For them*, reality and deception are indeed indistinguishable. Certainly, the birds left the picture of ZEUXIS quite soon: after being "disillusioned" as the food-detector was deactivated when the "food ingestion" reflex could not be performed successfully. Perhaps, a more adequate detector has even been activated instead. But the birds – conceived of as pre-object creatures – have no means of construing a relation between the two detectors, or the corresponding pre-objects respectively. By deactivating the food-detector, the intentional pre-object "food" just disappears from their world.

In consequence: if 'recognizing resemblance' is based on the ability to distinguish between 'being equal' and 'being seemingly but not really equal', then clearly, the birds cannot have seen the picture of ZEUXIS as resembling grapes. Because for them, it is not a question of 'being similar to grapes' but strictly of 'being equal to grapes' (i.e., rather just 'being grapes'). It might be prudent to distinguish in the discussion of images between two concepts of resemblance: concept »resemblance$_{\alpha}$« applies in all cases similar to those birds: a schematic classification is rated by some observer as erroneous. Concept »resemblance$_{\beta}$« is used if the creature we observe performing a classification does – or is at least in principle able to – understand that that classification is erroneous. Obviously, »resemblance$_{\alpha}$« cannot be conceived of as a relation derived as a weaker form of identity. It is logically more elementary and forms the basis for originally establishing the field of concepts of identity. In that latter field, »resemblance$_{\beta}$« takes the con-

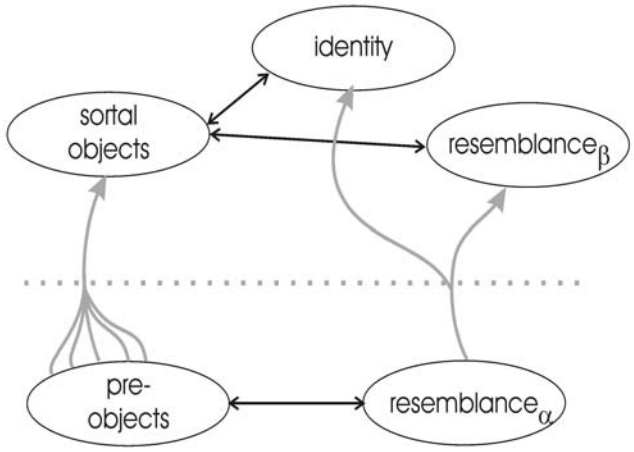


Figure 20: »resemblance$_\alpha$«, »resemblance$_\beta$« and »identity«

ceptual place of »resemblance$_\alpha$« (now in opposition to »identity«), thus opening the possibility of perceptoid signs. Though at its core is still the older primary resemblance only modified by the new possibility of individual objects.

For the complete object constitution necessary to understand how resemblance$_\beta$ is conceived, another transition to an even more complicated field of concepts of behavior must be considered: we have to look at creatures that can establish relations between pre-objects in *arbitrary* contexts, and thus originally invent identity in the proper sense (Fig. 20). The anticipations of pre-object creatures essentially integrate contexts *directly linked* by means of distance stimuli. That is, those contexts are organized along the course of coherent activity – they are, in an extended sense, all *present* in the course of the ongoing action. Hence such anticipations are not sufficient to establish identity between completely disparate situations. A re-*present*-ation of any contexts apart from those mediated by continuous action can only be mediated by means of signs (cf. also [PLESSNER *1928, Sect. 6.4*]).

3.4 Image and Language

The anchoring of indexical and iconic signs in language plays a major role for the second step of object constitution. Over and above the dependency of the concept »identity« from language use and the corresponding effects on our understanding of the relations between image and object, assertions like uttering the sentence "The photograph is blurred" are also imperative if we try to explain to what purpose a picture has been used. In particular, the interaction of the semiotic partial functions constituting language are quite well examined, and indeed help us to understand the corresponding interactions for picture uses. We continue by investigating the relation between images and assertions.

3.4.1 Assertions, Identity, and Contexts

Most language-analytic philosophers (e.g., [TUGENDHAT *1982*]) have reconstructed assertions as a specific composition of certain partial sign acts. By uttering "the photograph" in the example above, the speaker performs a *nomination*, with "is blurred" a *predication*. Both partial acts are "unsaturated" [FREGE *1892*]: it is not possible to use such a partial act by itself in order to perform a complete act of communication.

Nomination and predication are determined by their function. Although they are present in every single assertion they are not necessarily bound in a fixed manner to certain types of words or phrases. A nomination is the sign act used to direct someone's attention to the object of which something new (informative) is to be communicated; this object must be familiar to both (all) participating interlocutors – it must be part of their

common "discourse universe" .[16] If the actual situation of communication and the material objects "within" are meant, deictic expressions together with pointing gestures answer the purpose of nomination particularly well. But the function of nomination is not limited to those objects: we may of course utter expressions like 'the last unicorn', 'the Platonic Form of beauty' or 'Otto von Guericke', as part of an assertion, too, in order to point out an object of a corresponding discourse universe. They work as long as the discourse partners are able to distinguish the object meant from the other discourse objects in question. In contrast to the nomination, predication has no immediate representational aspect: it is the sign act used by the sender to inform or propose to the others that a certain custom of distinction – a concept – is relevant and applicable to the objects named. In the example above, 'being blurred' carries the predication. In every assertion, one predication (that may indicate a complex combination of concepts) and one or more nominations concur systematically.

Assertions are *context-relative*: if the corresponding discourse universe is unknown, an assertive sentence remains essentially incomprehensible. The nomination can only be performed effectively if it is clear which set of objects is at stake at all.[17] Objects are never given in isolation. We always speak of objects as something appearing as a figure in front of a background: they are part of a *context*. The expression 'context' is used in the following for indicating a finite structured set of intentional individualized entities, i.e., objects (as something known by somebody) with relations between them. More precisely, the relation between propositions and contexts is one of figure-ground to medium. A proposition offers a unique figure-ground distinction with the predication as figure on the ground of the objects known already and identified by the nominations. A medium offers the potential of figure-ground dichotomies, i.e., for many possible distinctions. Objects, while forming the background for predications, are thus seen as figure against other objects, as well.

Based on this introduction, many different types of contexts may be considered: for example, discourse universes are contexts shared by several creatures that communicate with each other. The situational context corresponds to what a single creature perceives as (individual) objects from its present environment. Other contexts are analogous to the situational one, but entail the "objective" environment of other times and places or even of fictitious and hypothetical situations.

Nomination can primarily be anchored in the overlapping parts of the interlocutors' situational contexts. The physical environment of the sign act (and all simultaneous behaviors of the interlocutors) provides then, it seems, the discourse universe of the objects being commonly perceived.[18] In contrast to that, the objects in the contexts evoked by previous assertions – or *co-texts* as they are called – do not have to be physically accessible. An earlier characterization (the concept used in predication) can be applied as part of a consecutive nomination in the form of a definite description: 'The blurred photo was taken by Hermione.'

[16] Proper names ('Harry Potter'), deictic particles ('this', 'she', 'you know who'), definite descriptions ('the picture of the fat lady'), and deictic descriptions ('this blurred photo') are the forms of nominations traditionally considered.

[17] Though the power of words for spontaneous context-evocation should never be underestimated, see below.

[18] Note that the different individual perspectives (as of pre-objects) must have been integrated conceptually in order to allow us of speaking from anything "being commonly perceived".

Objects (as we usually understand the expression) are members of many contexts. What we call the identity of objects is basically the question of connecting an object in different contexts. Take for example a court of law trying to establish the identity of, for example, the dagger now presented (1st context), the pointed object that was used to stab the victim one year ago on the other side of the city (2nd context), and the knife bought by the accused in the neighbor city 13 months ago (3rd context). Note that it is impossible to actually perceive simultaneously all the contexts in order to directly establish truth about identity. An important distinction of contexts – in FREGE's terminology (cf. below): of the "ways of being given" of objects – is the one between *referential* contexts and *intra-lexical* contexts. Objects are said to be referentially given if they are elements of the current situational context. In this case, the legitimacy of an assertion can be tested directly at the object by, coarsely speaking, including it in corresponding sensory-motor behaviors. The sensory-motor anchoring in the referential context is obviously the foundation of any empirical research. If, however, an object has only been introduced verbally in the discourse universe, there remains nothing but to apply conceptual rules and draw conclusions from the predication about the objects that are not explicitly mentioned, and to check whether the assertion is logically compatible with the context [SCHIRRA *1995*].

Assertions allow us apparently to make any context whatever a discourse universe, to share or harmonize it with the others, that is. Harmonizing perceptions between interlocutors by means of the sign acts has the obvious purpose of combining the diverse perspectives of an environment. Creatures thus endowed can perceive not only with their own senses, but also with the other's senses; they can manipulate not only with their own hands but with the other's hands, as well. Still, this would be a very restricted employment for assertions compared to what we usually do with them: humans mostly talk about objects that none of the interlocutors can actually perceive in the situation – or that may even be not perceivable at all.

That is, assertions allow us to relate an arbitrary context with the current situational one. The use of proper names given in a christening situation long ago depends on that ability. As was noted above, speaking of a deception – viewed as an explicit lack of identity – also means to relate two different contexts of behavior with each other; so does considering resemblance$_{\beta}$ (in particular with something being absent). Thus, being able to use resemblance as a crucial component of a certain type of signs (iconic/perceptoid signs) depends on a faculty that appears to be essentially mediated by assertions, disclosing a strong conceptual dependency between assertions and perceptoid signs.

In summary, assertions are context-relative on the one hand; but on the other hand, they are *context-independent*, since we can, at least in principle, perform an assertion relative to some context in *any* situational context whatever. The two characterizations of assertions depend on each other because it is only possible to speak independently of the *actual* situational context if another context can be explicitly referred to.

3.4.2 Communication Among Pre-Object Creatures

In a Gedankenexperiment, TUGENDHAT [*1982, Sec. 12*] mentions a simpler class of communication games that are not independent of the situation of utterance: the "quasi-predicates". A spontaneously sounded warning cry – 'FIRE!!' – may serve as an approximation of that type; but also an infant's utterance ('bow-wow!') in the one-word phase (ca. 20th month, [LOCK *1993*]) is closely connected with quasi-predicates. Their

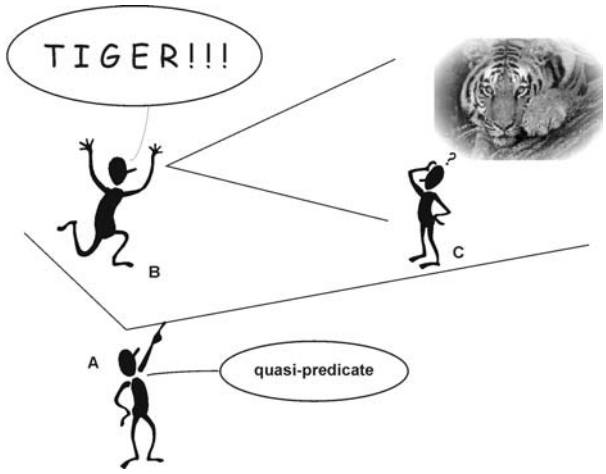


Figure 21: Ascribing a Quasi-Predication

rules of application must be strictly related to the corresponding context of utterance: a particular quasi-predicate is uttered only if certain conditions are being perceived (Fig. 21). The partner in communication also reacts on such an utterance (as a quasi-predicate) only directly, e.g., by taking flight or by calming the excited child. That is: judging the correct use of a quasi-predicate (or explaining the correct use[19]) is to be conceived of as being bound strictly to the situation of use. There is, so to speak, a fixed association between the (correct) use of a quasi-predicate and certain sensory-motoric routines (for testing or as a reaction).

Therefore, quasi-predicates articulate habits of distinguishing similar to a predication. The difference becomes prominent if we compare the situations of utterance of the warning cry 'Fire!' and of an assertion like 'The house is burning': the rules of use of the predication can be discussed in absence of a concrete example; situation of usage and situation of explanation can be separated. Correspondingly, the spontaneous reaction can be dispensed with in the case of a predication. That is impossible *per definitionem* when using the expression 'fire' (or also the sentence 'the house is burning') as a quasi-predicate. Quasi-predicates belong to the class of *signals*, a basic form of communication widespread among animals, and also part of the biological endowing of the anthropines (e.g., primary affective utterances; cf. [EIBL-EIBLESFELD *1984, Sect. 4.3*]).

Of course the understanding of a child's utterance 'bow-wow!' as the child telling us about a perceived individualized object suggests itself. But it is the field of concept of pre-object creatures that already allows us to construct a concept of communication based on quasi-predicates.[20] Let us assume for the moment that the repertoire of behavior of the child is (still) quite simple, so that we cannot yet speak of him/her perceiving objects in our usual pretentious sense; but we may speak without problems about detectors and the perception of pre-objects. We can "explain" to that child that he/she used 'bow-wow!' in the wrong manner or we may "confirm" the regular use; but only as long as that pre-object is perceived – i.e., is part of the situational pre-context.[21] Furthermore, we can provoke "accepting" reactions if we use 'bow-wow!' in the appropriate way: if the child perceives the corresponding pre-object; or the child begins to search for it. If there appears no corresponding perception the child reacts quite disconcerted. It would be rather odd if we react in a similar way on the nomination 'our neighbor's dog' in case it is not simultaneously part of our situational context. Thus, assuming the child is

[19] Note that such an explanation (in a wide sense) becomes necessary if the habits of distinguishing are no longer fixed genetically but by means of training, habituation, learning, etc.

[20] Evidently, the simpler creatures particular to the field of concepts of »reflexes« cannot be ascribed of having any sort of communication in a proper sense: there are only reactions on stimuli that *we* understand as being causally linked to other creatures.

[21] The 'pre' is necessary here since contexts have only been introduced on "full grown" objects.

a pre-object creature leads into interpreting its characterizing utterances as quasi-predicates.

Quasi-predicates are saturated: their communicative function cannot be set equal to neither the function of nominations nor that of predications. At best, the whole assertion may serve as an equivalent. But quasi-predicates *depend* on the situational context while assertions have become – despite their context-relativity – *independent* of the context of their utterance.

3.4.3 Context Builders and Referential Anchoring

In comparison to the elementary quasi-predicate, creatures gain with the complex sign act 'assertion' and its clearly separated partial acts 'nomination' and 'predication' the following essential advantage: it allows them originally to communicate about objects that are not present – not perceivable for them – in their actual environment. Using assertions is the only way at all to make "reachable" other contexts of action.

Indeed, we continuously use more or less consciously many verbal indications of contexts / discourse universes. Applying tempus is a typical example, since assertions about past or future affairs do precisely not refer to the present situational context as their proper discourse universe; the latter must be derived from the former. The grammatical modifications of the verb are quite an implicit indicator. Explicit specifications of location and time may also serve to reconstruct the context used as the discourse universe to be considered further on.

In order to adequately fathom this crucial aspect of assertions we assume another necessary partial sign act beside nomination and predication. GILLES FAUCONNIER, who is particularly interested in the linguistic potentials and consequences of such a proposal from a cognitive science perspective, uses the expression 'mental spaces' for contexts; correspondingly he speaks of 'space builders' – the verbal constructs that open up explicitly or implicitly contexts as the relevant discourse universes [FAUCONNIER *1985, 17*]. In the following, the expression 'context builder' is used analogously for characterizing the *partial sign act* that in the frame of an assertion allows the interlocutors to reconstruct the underlying context.

A special form of context building is the sequence of previous assertions, the co-text: the (intentional) objects introduced or modified there may easily be referred to again by means of definite descriptions that employ the distinction mentioned before. Thus, each continuous propositional text can also be conceived of as the complex context builder for subsequent assertions: "*In Tolstoy's ›War and Peace‹,* Platon is shot dead by a French soldier" (context builder in italics)

The distinction between the referential and the intra-lexical "way of being given" of objects has already been mentioned: only in the first cases, assertions about objects can be empirically checked – or as we shall say: can the assertion be *referentially anchored.* Context builders pointing out locations give us at hand a method of how to transform the context meant by an utterance into the situative context in which the referential anchoring could actually be performed. Spatio-temporal coordinates play an important role [TUGENDHAT *1982, Sect. 26II*]. Referentially anchoring an assertion then involves two steps: first, one has to know / recognize how the sensory-motor test routines linked to the nominations descriptive part (i.e., the one using distinctions formerly mentioned) are "prepared" – positioned, orientated, etc. (by transforming the context pointed out by the context builder into the situational context); second, one has to know how to actually perform the sensory-motor test routines for the newly communicated habit of dis-

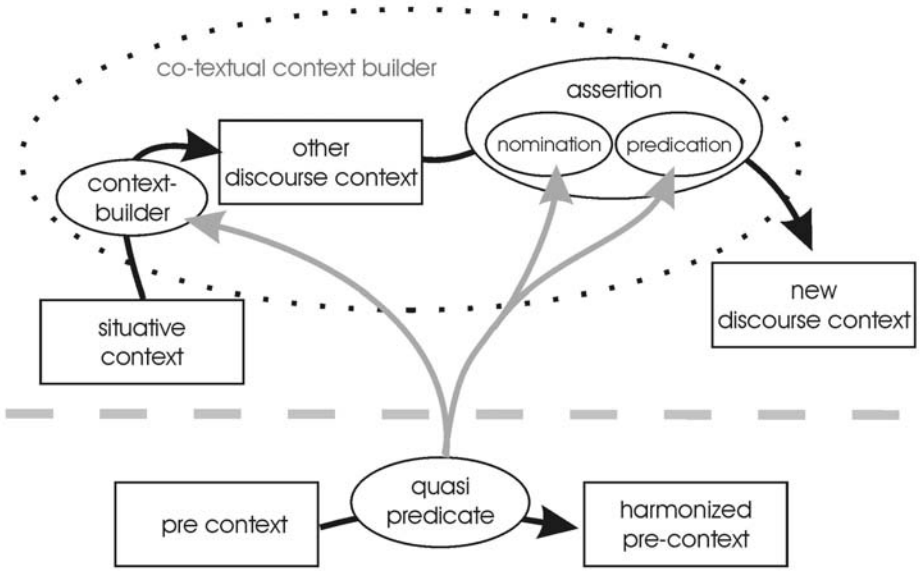


Figure 22: »Contexts«, »Context Builders«, »Nomination«, »Predication«, and »Quasi-Predicates«

tinction associated with the predication (e.g., that I have to *look* in order to recognize whether something – take the photo mentioned in the example above – is really "blurred"). Therewith, the "local" habits of distinguishing already associated with the elementary, i.e., strictly context-bound sign acts (quasi-predicates) may be employed. However, they are modified by extra conditions that are necessary for the individuation of objects, i.e., the integration of their absent aspects. Such conditions are essentially stabilized socially.

Assertions may be conceived of as derived from quasi-predicates (gray arrows in Fig. 22) since they fulfill a very similar overall function – to harmonize situations of behavior. The additional differentiation into the three clearly distinguishable partial acts "context building", "nomination", and "predication" is the precondition for redeeming communication from the strict binding to the actual situation. So far, all contexts but the actual situation of communication can apparently be constituted only by means of being verbally evoked. That is, we are in the interesting situation of considering, on the one hand, creatures that are able to communicate in an elementary manner but are in a way completely restricted to the "here and now".[22] On the other hand, we think of creatures with a more complicated behavior; they master a kind of communication that is independent of the actual situation. However, this art of a relative independence from situation depends circularly on their ability to communicate in such a complicated manner. The tool for overcoming that horizon is given only in communication. We hardly know yet how to understand this sharp transition (cf. [ROS *2005*]).

The problem we have reached here is indeed the question of the origin of (the field of concepts of) geometrical space (and measured time) *per se*: the medium needed for containing objects in the full-blown sense. A strange abstraction is necessary here for pre-object creatures: to learn to differentiate the places in space (and time) from the events and (pre-)objects "there". Perceptoid signs may play a crucial role for this step, though this is not the place to continue investigating this thread of thoughts.

[22] This includes more precisely all the locations in space and time that are directly connected with the present activity, i.e., not just a single (ideal) point of time or space.

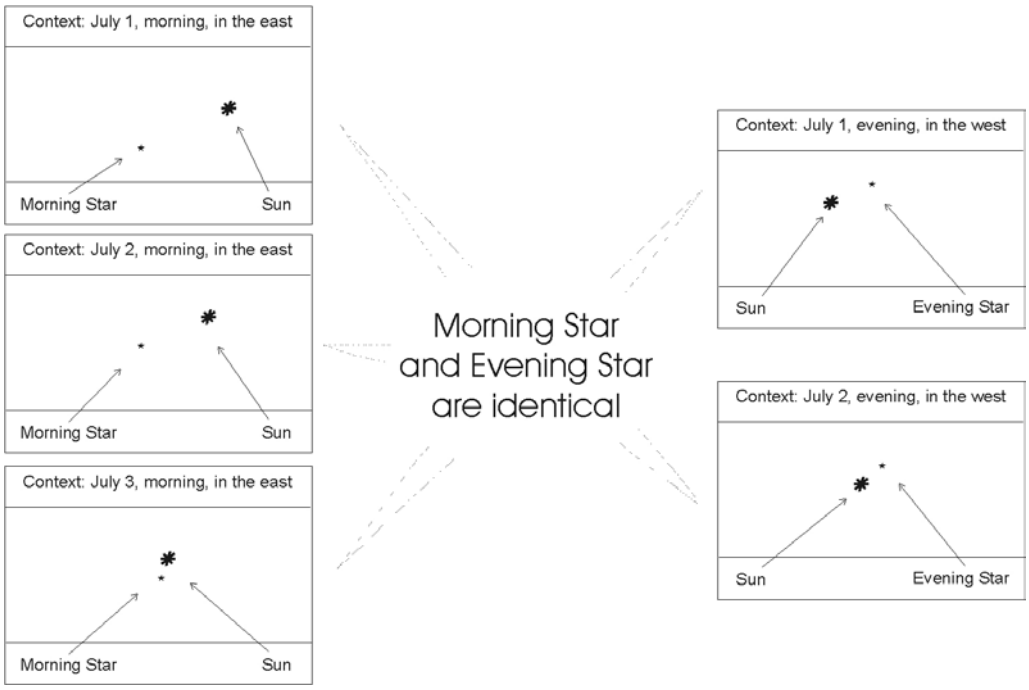


Figure 23: How Can the Morning Star Be Identical to the Evening Star?

The human fact par excellence *is perhaps not so much the creation of the tool but the domestication of time and space, i.e., the creation of a human time and a human space.*

[LEROI-GOURHAN *1984, 387*]

3.4.4 Secondary Object Constitution: Sortal Concepts & Geometry

While pre-objects are always referentially anchored, but cannot be accessed from another context, objects in the sense we usually associate with the expression 'object' are the constituting parts of many contexts. They are essentially viewed as instances of *sortal concepts*: perceptible, countable entities that are persistent over time even if they are not perceived, and that may even change their appearance dramatically during their lifetime (e.g., catapillar to butterfly). These are apparently also the kind of objects depicted in the most central cases of the concept »image« – from the animals of prehistorical cave paintings to ZEUXIS' apples and grapes, from the author's passport photograph to JUAN GRIS' cubistic portrait of PICASSO (Fig. 6). In his *Elements of Arithmetics* [FREGE *1884, §54*], FREGE distinguished this kind of concepts that "separate clearly and do not allow arbitrary divisions." A chair, by means of being a chair, clearly can be separated as an individual from any other chair; and the parts of that chair are not also chairs again. Objects falling under concepts like »water« or »red« do not have these attributes: two red objects are not distinguishable by means of their being red alone. And every part of a red surface is also red. Furthermore, sortal concepts allow us for pursuing an individual object in its singular temporal development across the contexts.

How can we be sure that something we saw this morning, e.g., a very bright star near the rising sun (let's call it 'the morning star'), and something we see right now in the evening, e.g., another bright star near the west horizon (correspondingly called 'the evening star'), are the same object? Or: what is actually the communicative function of an assertion stating the identity of the morning star and the evening star? It is in fact the

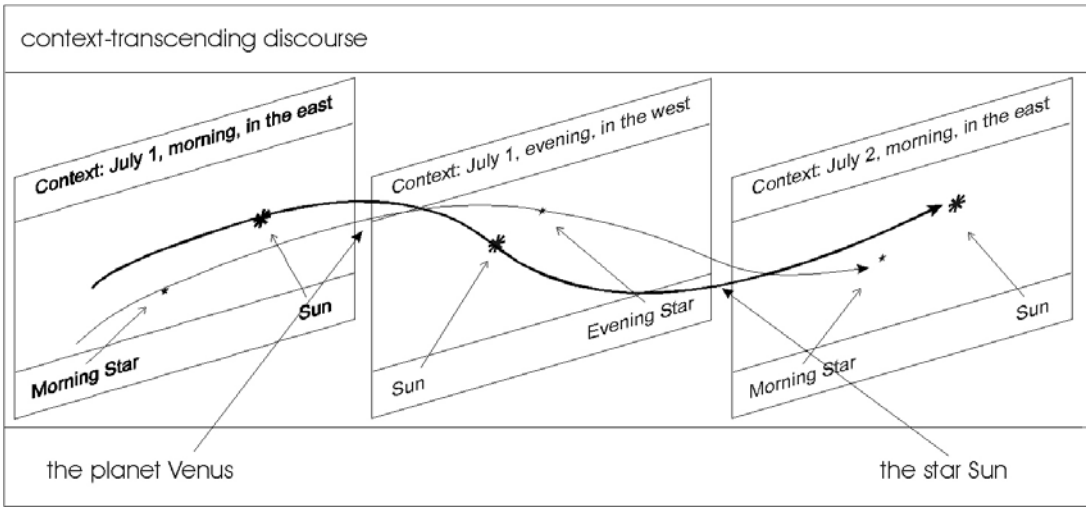


Figure 24: Identifying the Appearances of a Sortal Object (Planet) in Several Contexts

attribute of »planet« to be a sortal concept that renders an identification phrase like 'the morning star is the same as the evening star' to a meaningful utterance although the perceptual contexts of the two nominations are incompatible: the assertion has to be understood as 'they are both the same planet' (cf. Fig.s 23 & 24). That the referent of 'the morning star', which can merely be perceived in the morning, and the referent of 'the evening star', which correspondingly can be perceived only in the evening, are in fact identical, this proposition cannot be verified but by means of the criterion of spatio-temporal individuation given with the concept »planet«. In contrast to the planet Venus, i.e., the whole spatio-temporal extension of that object from its birth to its present existence (and beyond), which is only abstractly given and forms in FREGE's terms the common "Bedeutung" (reference) of the nominators 'the evening star' and 'the morning star', the immediate sensations of the Venus at either the early morning or the late evening, are the *"Gegebenheitsweisen"* – ways in which the Venus is presented to us, and also the only concrete way for it to be given [FREGE *1892*]. Obviously, these 'ways of being given' are closely related to pre-objects, their associated sensory-motor routines, and the referential anchoring basing any empirical observation.

Similarly, the referent of 'this house' while uttered from one particular point of view, and of 'this house' while uttered from a very different point of view may be the same house. In this case, the two nominators, which are in fact linguistically the same with the exception of their perceptual contexts, refer to two different manners of presentation of the same individual house meant – or, from the prespective of pre-object creatures, to two unrelated pre-objects.

In analytic philosophy, sortal concepts are conceived of as a systematic co-ordination between (a) configurational 'Gestalt' entities (of a 'geometrical field of concepts'), and (b) objects involved in part-whole relations that allow us to assign functions to those objects (of a 'functional field') (cf. Fig. 25; [VIEU *1991*]). The field of objects with the functional part-whole relations, abstract as it is, does not describe or restrict in any manner the geometrical relations between an object and its parts. It only allows us to state that there are such parts, and that without this or that part, the whole object would be something different. The schema of sortal objects leads to entities that have not only parts, but also a geometrical shape and a location; and additionally, all the parts *also* have shapes and locations – the whole object is a configuration of the shapes of its parts.

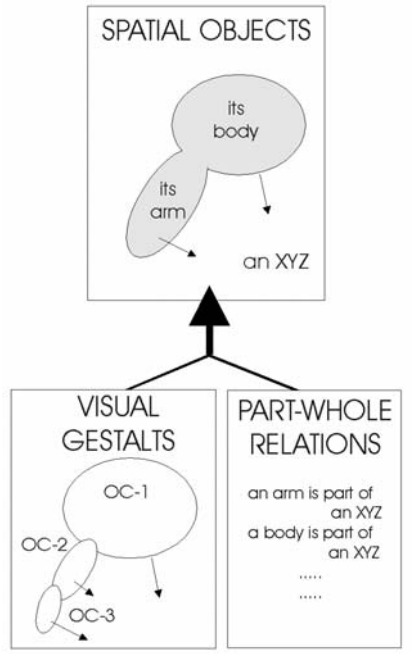


Figure 25: Sketch on Sortal Object Constitution

Note that the pre-objects or views at different "time slices" form another kind of parts of the whole sortal object.

The combination of the two fields of concepts has an interesting effect on the ability to identify corresponding instances: similar to two red objects, which are not distinguishable by their being red alone, the functional parts of a car, for example, do not distinguish one car clearly from another one of the same type, since they both have the same functional structure, and are therefore functionally indistinguishable. Only the different geometrical components of two instances of »car«, their different *histories*, allow us to distinguish both. It is, on the other hand, not the mere geometric Gestalt that makes something a car, but the functional restrictions between its parts. Furthermore, it is not possible to distinguish purely by geometrical features an object from its material, e.g., a ring and the gold making it up: in many contexts, these two different objects, which stand in a particular functional relation, have the same Gestalt properties and are involved in the very same geometric relations.

We remember: a pre-object is usually perceived by means of just one of the sensors of the reflex arcs combined in the corresponding detector. In analogy, a sortal object is perceived by means of just one detector: of all the pre-objects covered by the concept the one pre-object that is possible in the actual situation. We cannot perceive an instance of a sortal object in its whole spatio-temporal extension but only what is given in the one, present situational context. But whereas the concept »pre-object« does not include an option of accessing the constituting reflexes as opposed to the whole pre-object, it is a central feature of sortal concepts that the corresponding "manners of being given" can be made explicit: the current perception (as of a pre-object) in opposition to the individual with its complete history.[23] The ability to separate the different views is indeed equivalent with the ability to access other contexts (i.e., holding the non-current views of a sortal object). The integration of a multitude of contextual views also subsumes the co-relation of the distinct perspectives of the different interlocutors in one situation.[24]

The field of concepts of geometric Gestalts is of particular interest for us. The instances in this field correspond approximately to visual pre-objects. They are immediately observable. But they do not have the persistent identity of sortal objects and disappear if the beholder stops keeping them in his/her focus of attention. In contrast to mere pre-objects, they form however an incompatibility area of locations – a Euclidean coordinate system of potential, that is, not actually realized, situational contexts: space *per se*. Because that is after all what empty space is to be conceived of: as an infinite poten-

[23] – in opposition also to the abstract functional part-whole constituents, of course.

[24] For MEAD, the anticipation of the perspective of another one forms the crucial step for pre-object creatures to reach "significant gestures", i.e., context-independent communication by using signs with a common meaning, cf. Section 3.5.3.

tial of situational contexts together with a structure that allows us to reach one situation from another one.[25]

> *True individuality and true generality depend on the ability to consider the negative as such, the lack of something, the absence, the void. Homogenous intuition of space and time, hollow space and hollow time with empty places "needing" to be filled with constant elements are thus necessarily coextensive with true objective perception of things and true ideative abstraction.*
>
> [PLESSNER *1928, Sect. 6.5*]

As a consequence, if told the assertion that an object is at place *ABC* we do not – as usually with assertions – interpret the nomination (which object) and check whether the distinction mentioned by the predication holds or not (is at *ABC*), but go looking at *ABC* and then check whether the object is there (verification inversion).

3.4.5 Pictures as Context Builders: Resemblance Once More

In the previous sections, we have elaborated the semiotic partial functions of assertions to an extent that might be considered a bit exagerated in a discussion on pictures. Though, at least on first view, pictures seem also employable quite well for performing predications, and nominations, too. The standard function of a photograph in a passport, for example, can be understood as a predication: "This person *looks like this*". The photograph then activates a rather complex visual habit of distinction, which is linguistically integrated within the assertion's predicative part by means of the second use of the deictic particle 'this'. The nomination "this person" is implicitly clear in the situational context. On the other hand, it does not appear unusual if somebody presents the picture of a large red suspension bridge with two remarkably designed piles, and additionally tells us " 's been build in 1936." Here, the pictorial sign act takes over the role of a nomination. As is the case for purely verbal nominations, the passive discourse partner must consider that object as one already mutually known. In both cases, the pictorial sign act seems to receive one partial act of an assertion, the other partial acts of which remain unsaturated and unclear without the picture. The pictorial sign acts are not associated to any of the partial acts of the assertion *per se* or in any obvious way. The need for saturation of the verbal acts co-occuring with the presenting of the image originally induces the manner in which the picture is employed.

Nevertheless: in the light of the considerations about contexts and context-builders above, there is another possibility. Is it not quite tempting to understand pictures as fictitious referential contexts ? And in consequence: to interpret the communicative act performed by presenting the picture as a sign act of type "context building"? The presentation of the picture certainly enables the interlocutors to employ a discourse universe for their assertions different from the actual situational one. After all, the objects depicted are usually not also part of the latter. But we can use assertions with nominators for objects "perceived in the image" (and a complementary predication) without any problems. They are intentional objects, individuals that may, but need not, exist in any real situational context. Thus, pictures play on the one hand a role similar to a co-text, i.e., a sequence of assertions, by which an ensemble of objects has been introduced into the discourse universe so that some of their attributes and relations are explicitly fixed while others are implicitly inferred. All these attributes/relations may be employed for

[25] A more extended elaboration of this field in terms of computer science is given in Sect. 4.3.2

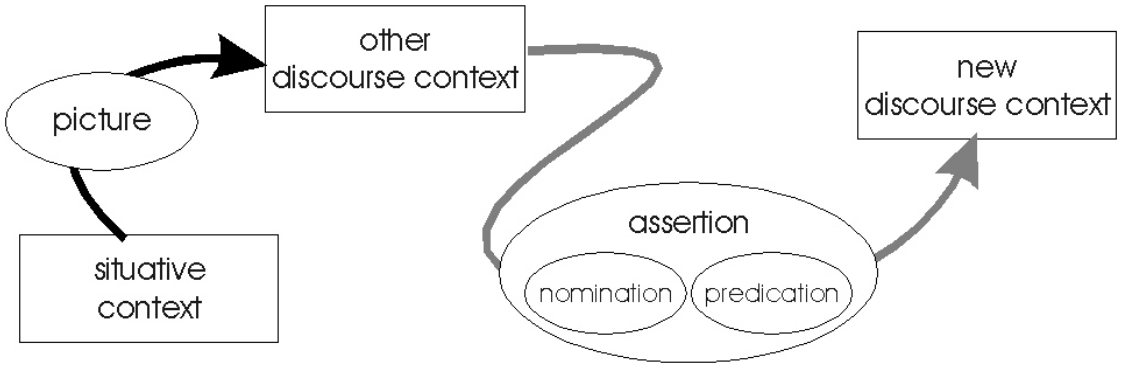


Figure 26: Pictures as Context Builders: Meaningful Only in Relation to Potential Assertions

the identification of the objects by a nomination in further assertions (compare Fig's 22 and 26). But on the other hand, and in contrast to the context building of a novel, the fictitious context opened up by a picture is *referential*: the objects are not again introduced by means of assertions (intra-lexically); they can be *perceived visually in the image* – in the case of the typical pictures mentioned earlier perceived almost as corresponding objects are perceived visually.

Therefore, the referential anchoring of the nomination (and hence: of a corresponding assertion) can be performed at least partially without intermediate steps: the according (i.e., basically visual) sensory-motor schemata apply to the image just as they do to a corresponding situational context. Context building by means of pictures may be characterized as a partial amalgamation of a context that is not present with the current situational context. This amalgamation of contexts may reach more or less continuously from a relative separation by "ALBERTI's window" to the complete immersion of virtual reality. Quite in analogy to the "usual" conditions of perception in the situational context, pictures are not integrated in a fixed association with certain assertions; they rather form an offer for interpretation, they open a potential for many reactions or interpretations in the form of assertions [FELLMANN *2000, 27 ff*]. These options are not completely arbitrary, restricted by means of the mechanisms of object constitution as they are.

Pictures are not employed primarily neither for nomitations nor for predications: those functions arise as secondary uses from the pictures' basic function as context builders. Let us have another look at the examples of pictures being used seemingly for nomination or predication mentioned above: the photograph in the passport, and the picture of the Golden Gate Bridge. Here, the picture's application as context builder is not only quite plausible. It is also compatible with the observation that the picture does not show the predicative or nominative function *per se* but receives them originally in its relations to the complete sign act and the other parts given explicitly. For the picture of the Golden Gate Bridge seemingly used as a nomination, this is fairly obvious: the picture opens up a referential context with a certain intentional object as a (potential) figur in front of a (potential) ground. Under the assumption that the interlocutor also recognizes the individual object visually, the year of construction is then verbally introduced as an additional concept holding of this object. Indeed, it is not the whole picture being used for nomination – the sky, the ocean, some vegetation or some ships though visible, too, are not meant in the example case. But they could be picked out occasionally as the relevant figures given by that picture, as well.

The presentation of the photograph in the context of showing the passport also works by making available another context: it complements the situational context, in which the (alleged) owner of the passport is present, with a second context, in which the real owner of the passport is semioticly present (at least visually) – that is, available for the corresponding sensory-motor test routines. The resemblance investigated is not one between (square) picture and (non-square) person but one between the appearances of the person at two different temporal contexts: earlier[26] and now. The assertion performed by means of the complete sign act "presenting the passport" comes out as a statement of identity between two "ways of being given" of the person, just like in FREGE's example of Morning Star and Evening Star. The picture's apparent function as a predication is therefore only derived from the more basic use as a context builder: the predication "to look like that or similarly" (a case of resemblance$_\beta$ taken as the symptom of identity) is something introduced originally by means of a "carrier object" *in* the picture (i.e., in that other context).

In conclusion, a first type of resemblance we called 'resemblance$_\alpha$' can be defined in the field of concepts of pre-object creatures where any concept of identity necessary for sortal objects is still missing. Remember that this type of resemblance was not recognized by the pre-object creatures themselves but only by someone thinking about them. The field of concepts that entails a concept of identity can only be conceived of as a field in which the differentiated interactions of the partial acts of assertions are established, as well. As a consequence of the ability to deal with more than one context simultaneously associated with that differentiation of semiotic behavior, resemblance appears in this field as »resemblance$_\beta$«– a similarity that can be recognized by those creatures themselves: they can be said to be able to distinguish between real and apparent.

The thesis has been proposed that a perceptoid sign act is a referential context builder – one of those partial acts constituting assertions – in a more or less restricted range of sense modalities: based on the merely schematic classifications associated with pre-object detectors in those modalities (i.e., resemblance$_\alpha$), we may introduce with the sign vehicle of a perceptoid sign something that for a pre-object creature could falsely activate detectors at least to some degree: the creature perceives an object – erroneously, but without being able to know that. More complex creatures however are able to cope with assertions, hence are able to understand sortal concepts and the relations between several situational contexts associated with them. They are thus used to distinguish between a sortal object and its context-dependent visual Gestalt, which may change while the identity of the object does not – or two of which may "resemble$_\beta$" each other for two quite different objects. They can employ the presented picture vehicle as the context builder part of an assertion, i.e., use it as a perceptoid sign.

Using perceptoid signs is, thus, unseverablely interwoven with the assertive sign system. But, in contrast to most other context builders, perceptoid signs allow the interlocutors for checking empirically the claimed statement at least partially by immediately anchoring the assertions referentially – that is by using corresponding detectors. The expression 'representation' as understood here is associated with the function of context building in general to bring the absent to presence, but may be used most clearly with the context amalgamation performed by means of perceptoid signs: the objects enclosed in the contexts communicated by means of a perceptoid sign can be reacted on verbally

[26] – more precisely, the time of registering the identity in order to construe the passports function: in a way a kind of formal christening situation.

and also (at least partially) in the more elementary and spontaneous bodily manner reserved for objects really present (i.e. as part of the situational context).

Note that as a consequence of this understanding it is impossible to conceive any image use for creatures that do not have the equivalent of assertive language. Without the separation of the functions of nomination and predication, awareness of a multitude of contexts cannot be established, conceiving sortal objects is out of the range, and the representation of any non-actual context remains impossible. Nor is there much plausibility for assuming a field of concepts of creatures that have only assertive language but cannot use perceptoid signs (not necessarily pictures, as there is no general necessity for a sense corresponding to the complex of human visual senses for creatures falling under that fields of concepts). The very idea of referentially anchoring assertions already includes the option for perceptoid signs.

3.5 Image and Image User

The previous review on analyzing similarity and the consequences for the semantics of perceptoid signs rests essentially on an act-theoretic basis: resemblance is not something independent from those recognizing or stating it; its concept corresponds in particular to the complexities of their potential behaviors. The relation between images and their users has therefore already formed a permanent background. In this section we rephrase and extend the results with an explicit focus on this relation and the pragmatics of pictorial sign acts. This also leads us to abstract, structural, and reflective pictures.

It has often been observed that pictorial signs seem to have a strange and ambivalent range of effects: on the one hand, their reception is eased by the mechanisms of object perception. On the other hand, this entices the beholder to uncritically interpret properties of the picture as properties of the object depicted. For gaining a better understanding of how perceptoid signs are used – and may be misused – it is of particular importance to reformulate their characteristics by bringing into the game different modes of reflection image users may take toward an object, sign or picture.

3.5.1 Reflection Modes of Dealing with Pictures

Let us first consider the *symbolic* mode of reflection, which is the usual mode we are in when using symbolic signs: somebody in that mode knows that the picture of a pipe is not a pipe, that he/she is employing something as a representative for something else, which more often than not is not present in the actual situation at all. As has been noted earlier, being in this mode also means to understand that a sign act takes place, which includes at least two participating roles and a sign that is part of a whole system of signs. The sign is a tool in or-

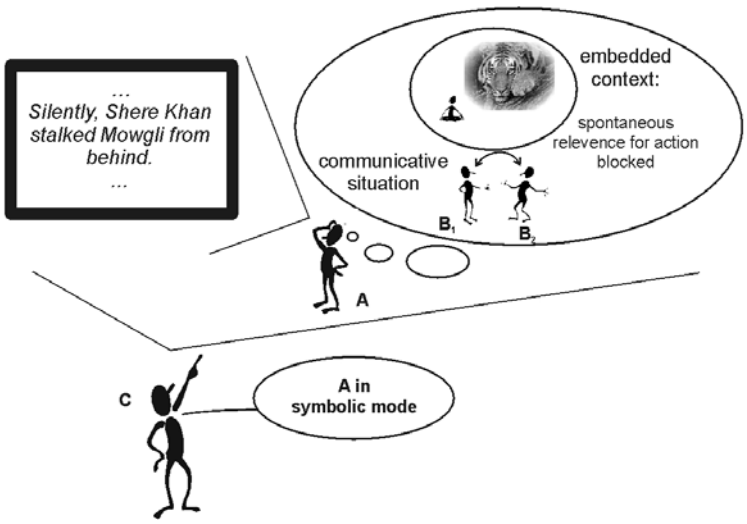


Figure 27: Ascribing the Symbolic Mode

der to coordinate the focus of interest of the participating interlocutors allowing them to act coherently together in a way that can even be negotiated and adjusted eventually. Independent of that task, there is just another object, the mere sign vehicle (cf. Fig. 27). Note in particular that it does not matter on this level of view whether the sign is a special kind of sign.

In contrast to the symbolic mode, the *deceptive* mode of reflection is given if we react on an object in a way as if a completely different object is indeed present. We may, for example, mistake the naturalistic portrait of a woman for that person – wondering perhaps about the strange stiffness she seemingly exhibits. This is of course not a case of communication at all – for somebody in deceptive mode toward a picture vehicle, there is no picture indeed. As should be clear, this mode is what pre-object creatures can reach at best. The deceptive mode is independent from the symbolic mode. However, ascribing the deceptive mode to someone (*A*) presupposes that the one who ascribes it (*C*) is him/herself in symbolic mode with respect to that picture vehicle (*V*): *C* must understand *V* as a sign for *O* in some communicative setting in order to see that *A*'s reactions fit to the presence of *O* but not of *V* (cf. Fig. 28). Note that the deceptive mode is exactly the stance we usually take with respect to ordinary perception: although we are quite aware in principle not only of the existence of sensory illusions, but also know about the complicated and not at all fault-free mechanisms underlying object constitution, we usually just trust our perception – we *are in* the world we perceive, a world consisting of sortal objects (among other more fluid things).

As we have seen in the previous section, the symbolic mode provides us with a virtual presence of objects and situations – in particular when dealing with assertions. Signs allow us to evoke other situative contexts, as is still preserved in the very root of the expression 'representation': "being brought back to be present". We can act correspondingly, e.g., feel a bit creepy when hearing the word 'snake'. But simultaneously, we must be able in symbolic mode to suppress most sensory-motor routines that otherwise would have fired "reflex-like". Or it would be quite hard indeed to finish reading a gothic novel. For example, the utterance of 'The Cologne city hall is burning!' gives us usually little panic compared with a signal call 'Fire!' (or similarly with the film of a fire in our room perceived in deceptive mode) – if only the Cologne council hall does not happen to be our actual situational context, of course. The context-evoking force inherent to the symbolic mode enables us to "put a distance" to the situational context and any spontaneous behavior inherently linked to it on the level of reflexes and pre-objects.

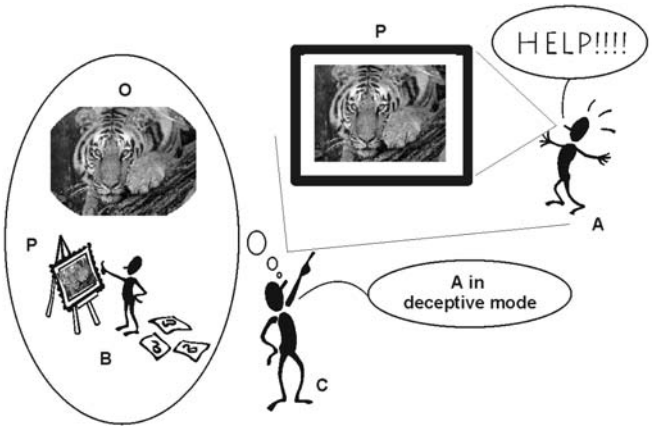


Figure 28: Ascribing the Deceptive Mode

There are very interesting mixed forms that appear particularly clear in the case of so-called virtual reality, but in a general sense cover all perceptoid signs: within virtual reality, a person acts at least partially as if a certain picture is indeed the object depicted, i.e., as if being in deceptive mode, though without losing the awareness of the semiotic foundation of the picture. This person acts, so to speak, consciously *as if* con-

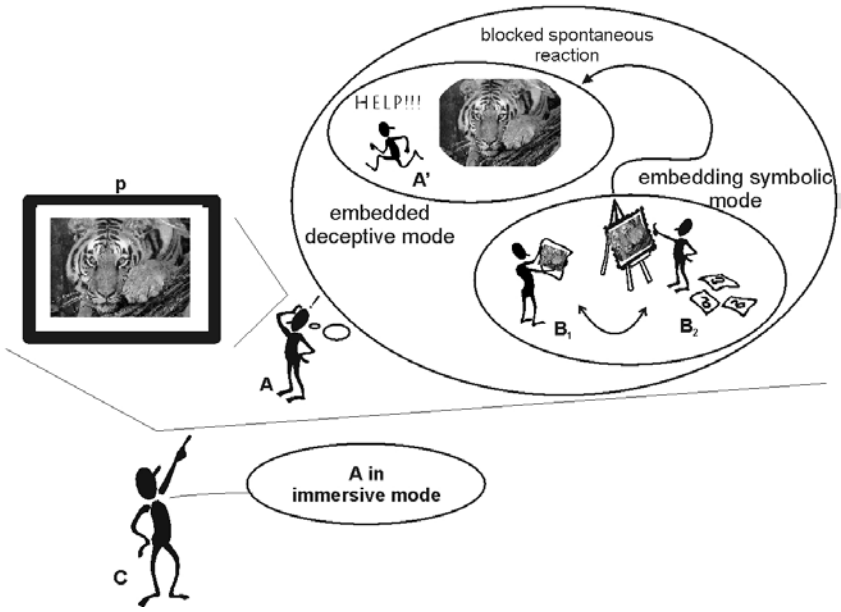


Figure 29: Ascribing the Immersive Mode

fronted with real objects. In contrast to the deceptive mode, this *immersive* mode of reflection, as we would like to call the basis of this acting '*as if*', presupposes the symbolic mode: we know the illusion, know that there is not really (let's say) a table, some chairs and a piano in the room we are in. But we actively engage with that illusion: for a creature in immersive mode, it is essential that the primary misclassification of the deceptive mode – a detector that is in fact wrong for the present situation, so to speak – is (and remains) activated, while the corresponding reactions can be suspended more or less like in the symbolic mode (Fig. 29). Resemblance$_{\beta}$ (as we have introduced the concept) is always involved in a pictorial sign act: it is only possible to conceive some things as "resembling$_{\beta}$" each other when being in symbolic mode while comparing the potential misclassifications for those two objects: that is, we have to be in the immersive mode in order to notice resemblance$_{\beta}$.

In order to ascribe the immersive mode, a correspondingly complex behavior must be observed: essentially (though this is not a complete characterization), the bodily reactions are mostly consistent with an interpretation of being in deceptive mode, while the utterances indicate the symbolic distancing.

The immersive mode governs the "ordinary" use of pictures, and doing so involves already a high amount of reflective competence. In theoretical discourse on pictures, an additional level of reflection is usually employed, which we shall call the *reflective mode*. Ascribing the immersive mode to somebody (*A*), as sketched in Figure 29, is possible only when being in reflective mode (*C*). As the functioning of pictures is to be investigated and explained in this mode, a focus on certain partial aspects is imminent: there are many examples of using pictures for directing our attention to one or the other aspect of picture uses in this book. Those example pictures are mostly employed outside there "normal" context of use, with special use conditions, similar to verbal examples given in a textbook on linguistics: it is crucial not to mix those special conditions valid for the reflective mode with the normal use conditions. We shall come back to the reflective mode in section 3.5.4.

3.5.2 The Game of Picture Making

Figure 3 – Reprise: Where does the picture end?

The previous paragraph focused mostly on the reception side of the communicative act, though for computational visualists the production side is rather prominent, too. Is a picture during the time of its production viewed always as a sign? Who is communicating then with whom anyway?

In fact, producing a picture means first of all producing a picture vehicle to be later used as a picture. There are of course cases where some object found may appear just ideal to serve the purpose, like a graining pattern of a piece of wood or even the clear surface of a lake. Though, recognizing/choosing the object as a potential perceptoid sign may be considered being equivalent to the explicit production of the sign vehicle in the other cases. Beside the behavior toward the picture vehicle, the picture makers must also be able to act with respect to what is symbolized – they must be able to show spontaneous reactions on the presence of that (sort of) thing – rudimentary or inadequate, as those reactions might be. Furthermore, they need to evoke a context of communication in general – they have to be able to anticipate the semiotic presentation of the picture vehicle by taking two different perspectives at once, as the sender and as the receiver of a message. That is, a picture maker has to shift perspectives a lot during the process of picture production, even if the purpose of the picture is decidedly to deceive a future beholder and not to communicate at all: taking the picture vehicle as a mere object with properties that can be changed in some way; evoking deceptive mode towards something else (i.e., evoking the spontaneous reactions together with the linked perception of something not present); evoking a communicative situation; and taking the picture vehicle as a sign used to "conjure" a fictitious context encapsulating the spontaneous deceptive mode for sender and receiver.

In the cases of naturalistic pictures, the visual appearances of sortal objects are usually to be primarily communicated. There are, of course, a number of different determinations of the meaning of the expressions 'realism' and 'naturalism' and their corresponding contraries depending on the context of discussion (e.g., in literature or epistemology). For our purpose the following conception has been helpful. Putting it simply, 'realism', as we understand the expression here, is the property of a representation of giving the impression of a configuration of spatial (mostly sortal) objects that is or could be found in the world. That is, fictitious objects are included whereas impossible objects are not (cf. Sect. 4.3.2.2). 'Naturalism' in our sense refers to the degree of a pictorial representation to which it evokes a visual impression as close as possible to that of the scene depicted. While realism is a binary category, naturalism only defines one pole of a continuous scale. The contrary to a realistic representation is one that either depicts non-spatial entities (e.g., air temperature, or the percentage of catholic households) or shows spatial entities as something "outside" the everyday space of three Euclidian dimensions (like pictograms of spatial objects in the abstract state space of an infogram). At the opposite pole to naturalism, a representation may still be realistic. But it does not use the "natural" visual impression of the spatial arrangement. Woodcuts, copper plate engravings or drawings with a pencil, even a black-and-white photograph give quite good examples of pictures that while being non-naturalistic are still realistic.

Total naturalism is a border case for realistic pictures that might make it difficult for observers to "see" the picture and not merely its content. Take for example again Figure 3: a quite extreme case for a live size *trompe l'œil* mural (ca. 3.3 x 8.6 m^2), the borders of which are barely noticeable at least in the pictorial reproduction given here and taken from the ideal viewpoint (cf. Fig. 30).[27] But at least its producer must view even a *trompe l'œil* in the immersive mode, i.e., as a sign in a communicative setting.

Normally, realistic pictures are composed of naturalistic and non-naturalistic elements. In a watercolor painting, the forms of the objects in a scene may lack naturalism while the colors are quite close to the visual impression of the "real" scene. A copper plate engraving may be highly adequate with respect to the depicted objects' forms, though we would rate it quite uncommon for those objects to show us nothing but uncolored crosshatched surfaces. Of course, it is the technique that restricts the modes of visual perception available for naturalism in these classical examples, and the producers of functional pictures often did not have much choice of technique in the past. As not only good "photo-realistic" computer graphic systems become increasingly available but also "non-photorealism" matures to a standard option in graphics systems that is quite diversified in form [STROTHOTTE & SCHLECHTWEG *2002*], designers of computer-generated pictures can already select quite freely between many techniques of representation with different aspects of naturalism.

"Truth" comes in mind as a good criterion of quality for both realism and naturalism of presentational pictures: truth of depiction of a real situation or truth of visual appearance. Take color photography: its indexical character seemingly guaranties both types of truth automatically, many believe. It has of course a precise sense to speak of an assertion as "being true" (or false, as may be the case): depending on the outcome of the sensory-motor test routines associated with the predication applied to the objects picked out by the nominations from the corresponding contexts, we may or may not agree with the assertion (or rather: the one stating it). Or, if it is not possible to referentially anchor the assertion, we check whether the new concept brought to mind by the predication leads to inconsistencies with what we know of that context so far. As it is the interaction of the two main components of an assertion – predication and nomination – that is originally responsible for ascribing truth, and in particular, for ascribing truth only *with respect to the given context*, we rather have doubts about conceiving truth as something directly applicable to perceptoid signs.

After all, only the assertions we are capable to generate from a given context can be associated with truth-values, not the context itself. Nor does the context builder qualify for truth-values. Pictorial sign acts are, then, not true or false. What can be used as a criterion of quality is at best the distribution of truth values assigned to the assertions generatable from the context built by the picture: compare it with the distribution of truth values assigned to the assertion producible with respect to some original situation. That criterion of quality is of little practical value due to the cardinality of the two sets of potential assertions, and of little theoretical value in the light of the context builder's defining function.

Though, closely related with truth is the question of authenticity. What is, for example, often meant when we say colloquially a film "is true" is that we rate the film as an authentic sign act. Remember that any sign act primarily shifts the interlocutor's focus

[27] A clear case of pure deceptive mode is hawed on the web page presenting JOHN PUGH, the artist of that mural, and his art: a patron of the café with that mural reportedly complained with the manager that the girl did not react on his trying so hard to flirt with her. (cf. www.illusion-art.com/incidents.html)

of attention toward the sender and his/her attitude, and only secondarily (if at all) toward what the sender's attitude may be directed at, an object, state of affair, etc. Authenticity is a general quality of sign acts indicating whether the sign act's primary focus, the indicated attitude, fits to the sender's actual attitude: not whether the sign act was true, but whether it was correctly performed, i.e., whether the sender was genuine, sincere with that sign act, and did not only pretend to be in that attitude. Remember in this context the tattoos and avatars mentioned in Section 3.1.

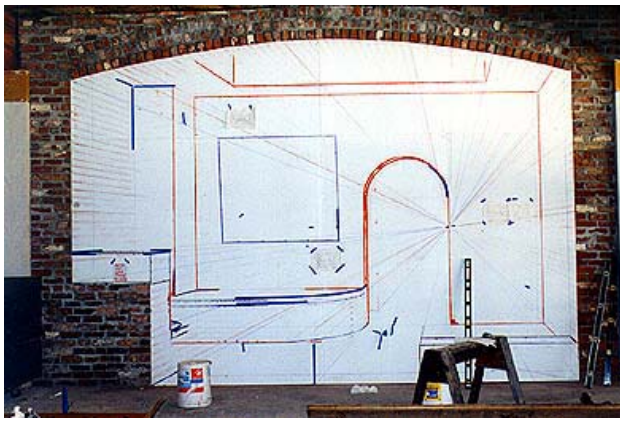


Figure 30: The Pictorial Answer to Figure 3

While truth is only applicable to assertions as a whole, authenticity can be ascribed to any partial sign act as well. One aspect only of this general concept »authenticity« is covered by the technical term 'authenticity': is the apparent sender of the message, e.g., a picture, its real sender/producer?

But what about the pictures occurring on the screen in a life transmission from a remote camera: is not »truth« the better concept there? And who is the sender, anyway?

3.5.3 Who Is Communicating with Whom?

In the preceding sections, perceptoid signs have been closely linked to verbal communication in the form of assertions. But we also find many pictures that are presented apparently free of any immediate relation to verbal language. Sure, the functional pictures we are mainly interested in appear seldom independent of co-occurring words. The use of sketches in maintenance instructions would be quite unclear without the explaining texts – just as the text remains incomprehensible if the discourse context the producer had in mind is not mediated by means of the sketches.

Furthermore, the presentation of pictures in a private photo album, too, has only sense if they evoke commentaries or explanations about the things, persons, and situations communicated by means of the pictures. Even if one browses the own photo album just alone, assertions form in one's mind that one could use at least as potential utterances for another party (*B*) – episodic stories or christening acts for persons who *B* does not know and about who one could further on talk occasionally to *B*.[28] The corresponding assertions need not be uttered aloud: it is sufficient to direct them in a kind of (inner) soliloquy to oneself while looking at the picture – or more precisely, while presenting the picture to oneself (in the role of an other person).

The pictures produced by means of the life transmission from a remote camera can also be interpreted as cases of perceptoid signs: it is indeed the beholder who directs his/her own focus of interest toward the situational context of the camera and the things there (or rather; who indicates to him/herself that his/her attitude is now to look at those things by means of the deceptive mode encapsulated in the symbolic attitude). Then, of

[28] This seems also to be the reason for browsing the family album of someone else alone without commentary being quite boring: the pictures in the album offer the possibility of stories, which are, however, not told. The contexts evoked remain unproductive, the partial sign acts unsaturated.

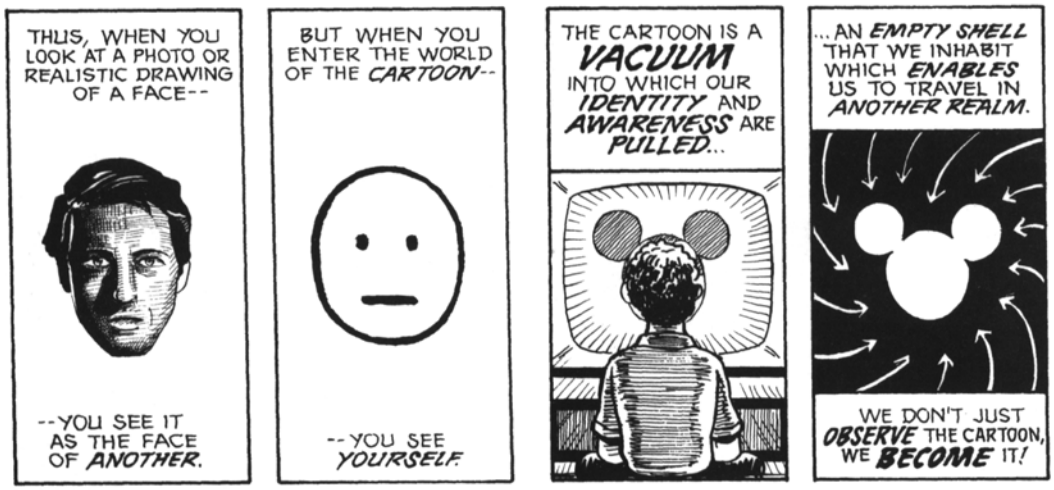


Figure 31: About MCCLOUD's Mask Theory of Faces

course, saying that those pictures are eventually "true" or "not true" has a derived meaning not too far away from that for assertions: when the self-directed partial act of context building (with respect to a non-fictitious context) is rated as being not authentic – e.g., because I as the sender know about a technical error of the camera or the transmission – then, of course, the assertions formed by me as the receiver on the basis of that context builder are not reliable with respect to that original context; but if the self-directed sign act of context building is rated as authentic, the assertions evoked for that context must have the same truth values as for the original context. Arguing for mirror images and other "natural pictures" as signs follows the same line of explanation.

Overcoming the restrictions of communication to the common situative context has been the dimension along which the three classical types of media mentioned already in Chapter 1 have been distinguished. With media of classes *II* and *III*, technical devices are involved so that a sender can communicate with a receiver at a totally different place and/or time. Usually, producing pictures is one such device (with the potential exception of ritually produced sand drawings mentioned earlier): the situation of production is completely different from the situation of reception. It is, then, of particular importance that the image producer keeps in mind the potential interlocutors and their possible situations of reception – and *vice versa*. If, as we have said, the beholder has to face the picture simultaneously in two roles, one of these roles is certainly the one of the probable (though possibly just imagined) image producer trying to communicate with us.

This anticipation of the communicative partners corresponds to G. H. MEAD's analysis of a crucial aspect of conscious communication distinguishing it from more elementary forms of communication like signals: in the sending individual, the same reaction is systematically triggered as is in the receiving individuals [MEAD *1968, 68ff*]. This concept of the role »sender« entails that any instance of »sender« has to adopt the role of the »receiver» in the sign act (and, in fact, *vice versa*). In order to be language – i.e., communication in an advanced sense – what is communicated has to be understood equally by *all* the interlocutors of the exchange. More precisely: the sender must in principle be influenced by his/her sign act in the same way as the others. For verbal language, such a conception is essential since speakers never mention explicitly everything actually communicated: the phenomena of ellipses and anaphora, presuppositions and

conversational implicature are just the tips of the iceberg. They can hardly be conceived without a speaker who is able to anticipate systematically the part of her/his listeners.

In another form, this is true for pictures, as well. We may, for example, interpret the lack of visual features in line drawings as a form of pictorial ellipsis. Even for *trompe l'œil*s – to be seen as such, and not reacted upon in mere deceptive mode – the close similarity, which evokes "correctly" the "wrong" detectors spontaneously and without a special phase of training in advance, leads only to understand *that* a deception takes place and may have been intended, but not to understand *why* this particular deception was placed there and then. That question concerns the communicative purposes of the picture use, and hence the two roles participating, together with the mutual anticipations of each other.

Figure 32: A Caricature

Even on the level of quasi-predicates, the perception of faces is important for communicative processes among anthropines. Faces are primarily pre-objects (or, on the complex level, sortal objects), but if we follow MEAD's analysis of communication they must be more than just that. Pictorial presentations of faces invite, so to speak, the beholder to identify the role of sender with them, in particular if eye contact seems possible. In his considerations on comics, SCOTT M^{C}CLOUD [*1993, 36*] argues that the offer for partial identification is indeed the reason for reducing naturalism and increasing the amount of abstraction (Fig. 31). It is much easier to identify with a figure without highly individualized facial features, easier to put on that mask, and to "walk" with it through the pictorial space (of comics, in that context).

3.5.4 Indirect Resemblances & Rhetoric Derivations

The final sections of the overview on the basics of image science are dedicated to more complicated manners of using pictures. Not all pictures represent sortal objects in a momentaneous spatial configuration. Beside those representational images, which may be more or less abstract, SACHS-HOMBACH [*2002, 145ff*] puts two other classes: structural pictures like diagrams, and reflective pictures. The latter appear often in art. They are called 'reflective' as they are used to communicate pictorially about the conditions of picture uses and picture productions, or for short: about picture communication and its constituents itself. Especially modern art has contributed many different aspects to that pictorial meta-discourse.

Structural pictures are (usually) not intended as a reflection on the uses and conditions of pictorial sign acts. They are part of the class of non-realistic pictures, i.e., they do not depict a spatial arrangement of sortal objects (with potentially a few non-sortal accessories), and it is usually not straightforward to understand how resemblance is involved in their interpretation.

Figure 33: A Rather Famous Picture with Obscured Resemblance Due to Color Reduction

Note that already realistic pictures in a non-naturalistic style, i.e., abstracted representations, call for a modification of the strict resemblance criterion. The sketchy line drawing of caricature may still spontaneously evoke a strong deceptive mode on the basis of a rather reduced set of prominent visual features (cf. Fig. 32). Other abstracted presentations need much more intellectual effort to activate the corresponding detectors necessary for recognizing what is depicted (Fig. 33). When we take resemblance as a general criterion to characterize pictorial signs, we presuppose that properties are ignored that do not contribute to the content of the sign and are irrelevant to its interpretation.

The fact, that resemblance is determined only in certain respects allows us to accommodate it to quite different pictorial phenomena. In some cases – mainly naturalistic pictures like photographs – we immediately and involuntarily regard most respects as relevant that are also relevant in perceiving objects visually. In others, like line drawings, we leave aside many of those respects: the picture is taken to resemble some object only relative to the remaining respects. As an extreme case, a diagram, for example, does not show us anything about how physical objects look like. It is possible to establish different respects as dominant because the picture vehicle does not in itself determine which properties are relevant for the depiction. Sometimes, relevant respects are even missing (Fig. 33). We may develop different pictorial schemata with respect to some particular communicative functions the pictures are supposed to perform. In this sense it is then plausible to say, SACHS-HOMBACH argues, that all presentational pictures resemble their objects *in one way or the other*: what resemblance exactly is to be used relays completely on the pictorial schema determining in each case which are the relevant respects.

Accordingly, understanding pictures involves two components: the image users have to decide what properties of all are intended as relevant for a perceptoid sign before they can integrate any spontaneous deceptive reactions in the immersive mode, and thus determine which objects it is that resembles that sign in those respects. The more the picture producer restricts the set of relevant respects, the more the picture users must know about the semiotic rules of that particular pictorial subsystem in order to be able to interpret a corresponding "abstract" picture adequately. Spontaneous deceptive reactions alone may become too weak. However, the greater the number of essential properties a picture has in common with another object (by means of resemblance$_{\alpha}$), the more easily a picture user recognizes that object in the picture in spontaneous deceptive mode. Looking at the picture then becomes more and more like looking at the object depicted itself, and the picture's style more and more naturalistic.

Abstracting in general can be understood as the process by which an extract of all the information available for some theme or scenario is refined so as to reflect the importance of certain aspects for the communicative situation at hand [STROTHOTTE *ET AL.*

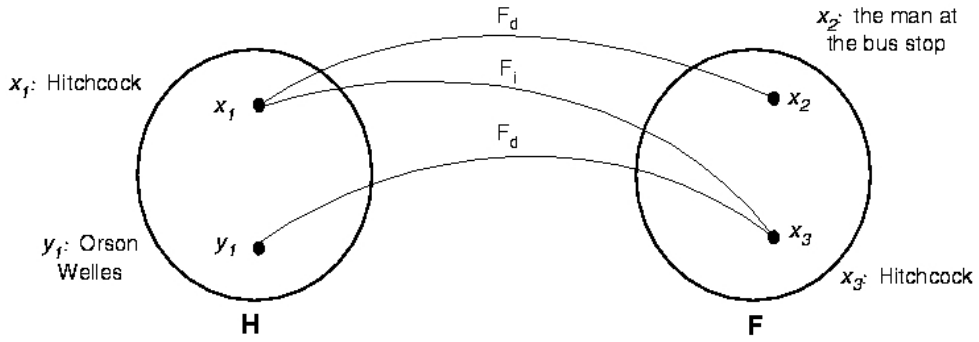


Figure 34: An Example from [FAUCONNIER *1985, 36*] – Metonymic Connectors between Mental Spaces (Contexts) as the Basis for Metonymy

1998]. It is especially linked with the rhetoric figures of metonymy and synecdoche. 'Metonymy', from Greek '*metonymia*' = renaming, refers to the rhetorical use of a word or expression for something that is closely connected in a spatial, causal or logical way to the literal meaning. 'Synecdoche', from the Greek words '*syn*' = 'together', and '*ekdoche*' = 'taking over', 'conceptualization', is often viewed as a central case of metonymy. Its more familiar Latin equivalent, *pars pro toto*, indicates that the expression for a part of the whole is used to name the whole. Typical examples in everyday life are evoked by: '*The White House* has officially denied any involvement in the murder', 'I'm *the ham sandwich*, *the quiche* is my friend', '*Wall Street* is in a panic,' and '*September 11* has severely changed the Western social climate'.

GILLES FAUCONNIER – mentioned already earlier as giving the motivation for the name 'context builder' by his studies on mental spaces and linguistic space builders – has indeed explained metonymies as a special kind of linking two mental spaces (his version of contexts). As an example, Figure 34 sketches two mental spaces relevant in the following situation: Suppose, a film is made about the life of Alfred Hitchcock, whose role is played by Orson Wells, while Hitchcock himself appears in a minor role as "the man at the bus stop". The context of the stars of Hollywood (H) is opposed to the context established by the film (F). The objects in the contexts are linked by the "drama connector" F_d (who acts as who), and by the "image connector" F_i (who is pictorially represented by who). Those relations are used to construct metonymic expressions, i.e., by employing an expression correct for one context to refer to the corresponding object in the other one: "In the third scene, Hitchcock was seen following Hitchcock" or "You mean, the man at the bus stop is played by Wells?"

The linguistic phenomenon of metonymy applies essentially to nominations, and with respect to this function, the transfer of the concept to pictorial sign acts must be viewed: a context builder that is used to introduce a context with objects standing in metonymic relation to some other objects actually meant to be contained in the context. In fact, the perceptoid sign acts always depend on a metonymic relation – taking the momentary appearance of a sortal for the temporally extended whole. On this basis, further metonymic abstractions can be easily integrated into the conception of pictorial signs.

An important concept in the context of abstraction is »exemplification«: we can use a concrete object metonymically as an example case for any concept that holds of that object in order to speak (symbolic mode) about this concept, e.g., a certain horse in order to discuss horses in general, a certain keyboard in order to debate keyboards in the es-

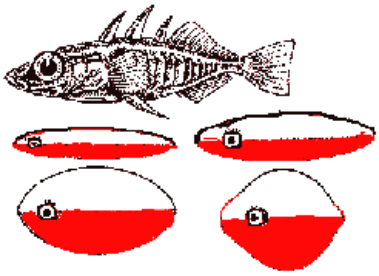


Figure 35: Supernormal Stimulus: The shape of the three-spined stickleback (*Gasterosteus aculeatus)* and several dummies

sence. The object then is used as an arbitrary sign[29] for the concept. Correspondingly, we may employ a picture as a context for introducing an object not as a particular individual but as an exemplification for some concept, in particular concerning the visual Gestalt constituents of sortal objects. This is often covered by distinguishing 'a picture of an X' from 'an X-picture' [SCHOLZ *1991, 26ff*].

Concentration of features – as in the case of caricature – is one means of abstraction that can even be linked back down to the reflexes integrated in the detectors for the corresponding pre-objects. The phenomenon of "supernormal stimulus" is well known among ethologist: for example, male sticklebacks behave in a specific way when in reproduction mood if another male stickleback is present ("defense of district" behavior). Due to the merely schematic classification possible with reflexes, the sticklebacks react already on dummies with only a coarse similarity (from our point of view, of course): elliptic shapes of about the right size trigger the fighting behavior if only the red belly spot typical for male sticklebacks in reproduction mood is present. The reaction on presenting the "abstracted" dummies is often even more distinct and stronger – they are supernormal stimuli for the sticklebacks (Fig. 35). We may take it as an educated guess at least that caricature is based on that mechanism, as well. The suggestive indicators also weaken in these cases the tendency mentioned above to identify with the figure.

Another source for abstraction can be reconstructed on the level of detectors. Let us assume that certain indicators may be among the distance stimuli integrated into the detector for one kind of pre-object that are not directly "emanating" from that (sortal) object: the fresh foot prints in wet sand or mud can be used as quite valuable distance stimuli for the animals having produced them, as well. For a pre-object creature with a "foot print reflex", the firing of that sensor activates the corresponding detector, i.e., it perceives the pre-object associated, and reacts accordingly – quite reasonable under many circumstances (e.g., foot prints of an enemy).

When continuing the argument on the level of sortal objects, the footprints can be viewed as one of the many forms of visual appearance of that object – one of its manners of being given visually. Consequently, it can establish a resemblance$_\beta$ relationship and be used as part of a perceptoid sign. Figure 36 shows an example of foot prints of several species in a *jukurrpa* picture, where the foot prints of mythical ancestors in human and animal shape, or even

Figure 36: Traces and Paths – Section of the Australian Aborigine Picture *Karrku jukurrpa*

[29] The object is not linked by causality nor by similarity to the concept, hence it is not an index nor an icon, and must be symbolic in consequence.

the U-shaped depression left by a sitting human (with the elliptic bump from the traveling basket beside) are quite prominent. Note that by means of the traces, sortal objects are in a way shown in an integral manner: the appearances at many moments are bound together. The pictures of Australian aborigine art are assumedly simplified or obstructed versions of the secret ritual sand pictures employed when the myths are enacted. Most interestingly, the traces are used to structure the myth's narration ([WATSON *1999*] and [RUMSEY *2001*]) – another indication to the strong conceptual connection with assertions by means of context building.

Figure 37: A Simple Hand-Drawn Map

Maps can be conceived of as directly associated to trace pictures, but form also a link to structural pictures. Maps can but do not have to preserve the full set of geometric relations, i.e., being "true" with respect to directions and distances. Distances may be not a relevant respect of resemblance (Fig. 37), or topology becomes the only interesting feature to be communicated (Fig. 38). Maps for the users of public transportation systems share this characteristics with state diagrams for Finite State Machines – a formalism often employed to describe simple programs, the behavior of non-player characters in computer games, or any system of actions to be performed in various temporal orders depending on some contextual conditions (Fig. 39). All these examples also abandon usually depth in the representation as a relevant aspect, though there are structural pictures employing the three dimensions of pictorial space in a non-literal way, as well (Fig. 40).

The relation between representational pictures with traces and diagrams has been mentioned already at the beginning of this chapter. The clue to that link is given by another rhetorical figure, namely metaphor, which may indeed help us to understand the phenomenon of structural pictures in general. 'Metaphor", from the Greek expression '*metapherein*' = 'to transfer' indicates the non-literal use of a word or expression based on an elliptic comparison: if Julia is for him *like* the sun in some respect, Romeo may say 'she shines so bright', or that the world is so dark without her. As a rhetorical means, metaphor seems to appear essentially in high literature, in poems, etc., but not as frequently in everyday language use: though think about "time is money", "life is a journey", and "movement is a path." In fact, metaphors transfer not single expressions but significant parts of a whole field of concepts into the other field, and thus allow us to generate more than a singe metaphor: a complete and coherent system of metaphorical expres-

Figure 38: A (Mainly) Topological Map

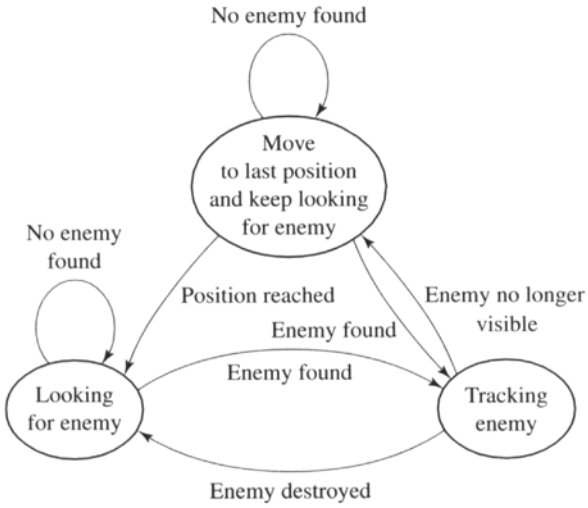


Figure 39: State Transition Diagram of a Finite State Machine – Actions in a Computer Game

sions. Time, for example, may be conceived as "flowing" like a stream of water, sometimes fast, sometimes slowly, sometimes troubled, sometimes calm; events drift toward us, meet or even hit us or pass by, etc.

While metonymy interacts with nomination, metaphors are connected with predication. Metaphoric transfers are often used to originally structure a new domain of experiences by means of a conceptual structure that is well-known and seems to be similar in some respect with what is yet known about the new field (cf. [LAKOFF & JOHNSON *1980*] for a general discussion of metaphor).[30]

The application to context builders is indirect again, but explains the foundation of structural pictures. It is the structure of the geometrical Gestalts that becomes the source of metaphoric transfer for a field of concepts not directly linked: time "is" space; concepts (or other abstract entities) "are" boxes, relations become paths (arcs or arrows). Set inclusion becomes geometric inclusion (note the linguistic metaphor here), multiple memberships occur as intersections; quantities "are" heights, etc. By means of that transfer, the elements of the goal domains firstly become something visible at all. Resemblance, then, plays a role for structural pictures, as well. But in their case, it is an abstract kind of resemblance, the same kind involved in metaphor in general: a partial isomorphism between fields of concepts. In contrast to sortal objects and their appearance in presentational pictures, there is usually no resemblance$_{\alpha}$ at the foundation since the abstract entities do not qualify for pre-objects *per definitionem*. The fact that words are usually integrated in structural pictures is another indicator for such pictures being based by metaphors: the connection to the target domain must be established in some way: "Julia is the sun" or "this circle is an option".

In Figure 41, several metaphors (and metonymies) are integrated with standardized pictograms and words connecting the graphical source domain of the overall metaphor with the target domain, the structure of a university degree programme. "Building" a degree by starting from the base and step by step finish-

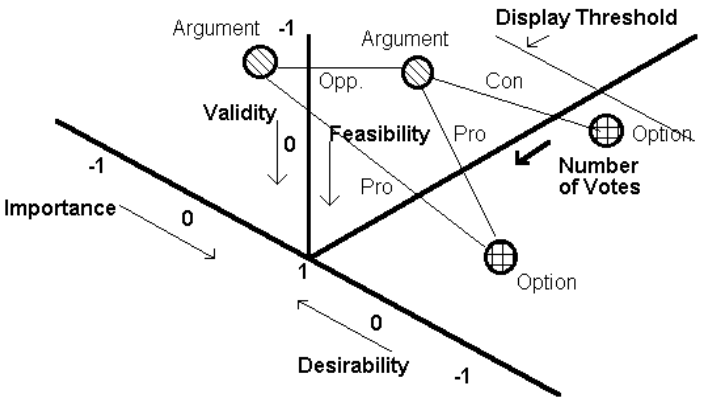


Figure 40: Metaphorical Transfer of Three-dimensional Space (visualizing a complex relation in election theory)

[30] Due to this use of the concept »similarity« in explanations of »metaphor«, the latter have sometimes been called "verbal images," "pictorial expressions".

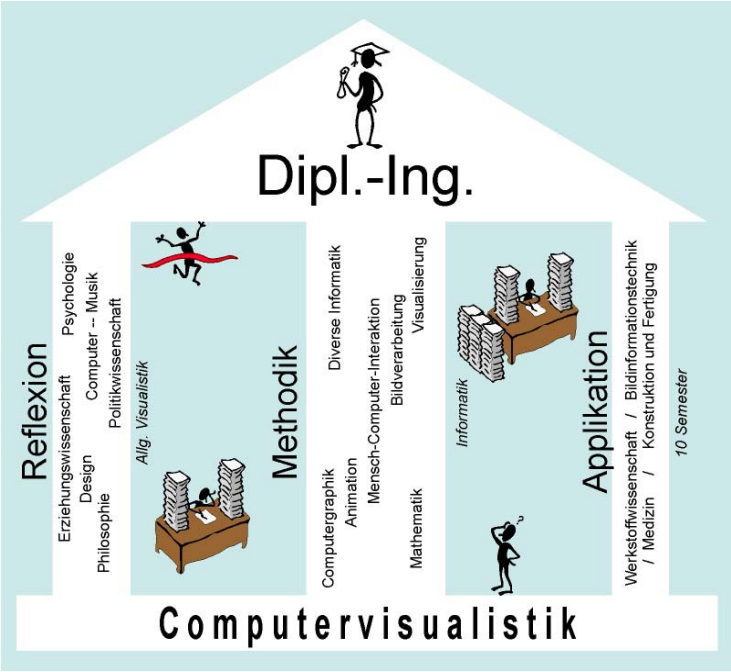


Figure 41: "The Three Columns of Computational Visualistics"

Pictorial Presentation of the Structure of a Degree Programme

ing the roof – a lot of work; "passing" an institution on a certain path; an institution structured by building blocks (like columns). This conceptual structures can be used as a "narrative map" , either by an explicit sender, or by the beholder, presenting the graphic to himself/herself and explaining in an inner monologue about that degree programme by traveling "through" the picture.

3.5.5 Reflective Communication & Pictures of Art

A visitor to a museum looking at a pictorial exhibit (of the representational kind) can be conceived of as presenting the exhibit to him/herself: s/he "tells him/herself internally" (so to speak) what s/he could tell somebody else about the scene depicted. This also forms the very basis for any stylistic considerations and aesthetic judgments (a path which we shall not pursuit here). In those cases, the quality of the constitution relation plays a particularly important role, i.e., the relation between sortal objects one thinks to see, and the visual Gestalts that the picture shows and that constitute the sortals. Many pictures of art are reflective pictures: as has already been mentioned in Section 3.5.1, they must be associated with a special mode of reflection, as the communicative act they are used for deals with the pictorial sign act itself, and hence with the immersive mode and its complicated inner structure. This may reach from exemplifying the ability of the picture maker to produce highly deceptive pictures (as plays a major role for many *nature morte* of the 16th century) to the pictorial critique of the focus on naturalism. The central theme of the American art style 'photorealism' of the 1960s and '70s, for example, is an indirect critique of the visual access to reality in the modern industrial societies: an access that is almost totally mediated by technical reproductions, and thus open to all kinds of hidden manipulations [HELD *1975*]. The images of artists like CLOSE, BELL, and MORLEY do not try to show reality in a photo-like realism; their

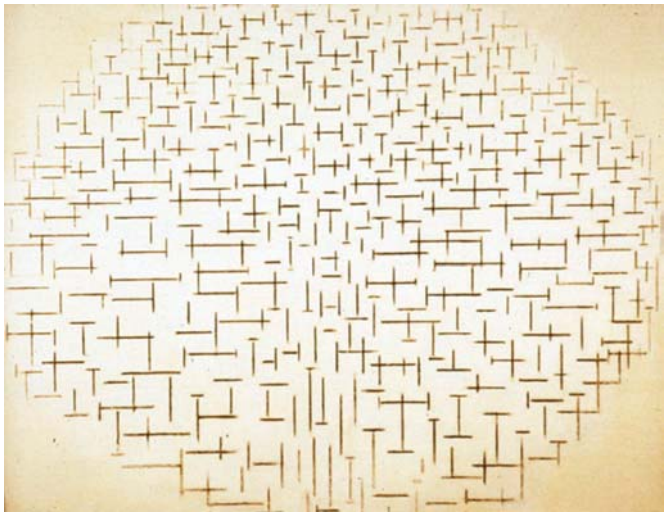


Figure 42: *Ocean and Pier*

P. MONDRIAN *1915*

subject is the mediated access to what is believed to be reality by media that are assumed to present subjects naturalistically.

The emphasis in cubistic pictures on the integration of many perspectives necessary for pictorial presentations of sortals has been mentioned already in Section 3.1 (cf. Fig. 6): such an integration forms, as we have seen, the basis of object constitution. The topic of cubistic pictures over and above the still realistic spatial arrangement of sortal objects is therefore, in a way, the difference in temporal quality between momentary visual Gestalt and persistent sortal object. Even further along this path, PIET MONDRIAN's famous paintings at the border of figurative and non-figurative pictures have evoked commentaries like: "The rapt quality of the image seems to embody a longing to deny time" [SYLVESTER *1997*] (Fig. 42). Other branches of visual arts concentrate on the materiality of the picture vehicle, in particular the screen and the pigment.

Although reflective pictures of the kinds used and invented in modern art are seldom relevant in computational visualistics, at least the particular use conditions of example images employed in texts on pictures may be considered important. We may quote pictures in order to exemplify a certain algorithm of image processing or computer graphics. Again, an aspect of picture production (hence use) is communicated by means of the presentation of such a picture; what is to be seen (as those pictures are usually of the representational kind) is more or less contingent. The frequency of teapots in pictures presented in computer graphics books does by no means communicate a particular addiction to the beverage (Fig. 43), nor does the insistence on skeletal feet in publications from the Magdeburg computational visualistics group indicate a very strange fetishism. How the object chosen is depicted, how the visual Gestalt relates to the sortal object, and in particular: how that relation again is linked with some aspects of the algorithm exemplified, that is what the sender of such a message normally intends – and what the receivers expect to be told in those communicative circumstances. Those pictures are therefore clear cases of reflective pictures, as well.

3.6 Conclusions for Computational Visualistics

The overview on image science given here was centered around the following idea: to isolate the essential properties of the concept »picture« by means of a logical reconstruction ("an implementation" see Sect. 2.1) of the corresponding field of concepts. Such an investigation deals with questions like "how can we motivate that such a form of sign acts with the special involvement of (visual) perception could have developed?" – "What are the general properties of sign use and perception that are necessary for such a conceptual combination?" – "To what purpose and under what

preconditions did creatures evolve that can use such a type of signs?" Thus, we deal with the concepts »perception« or »sign use« or »picture« *in the essence*, so to speak, not with a particular form any such concept may take in a certain cultural environment; with the "conditions of their possibility" in KANT's words.

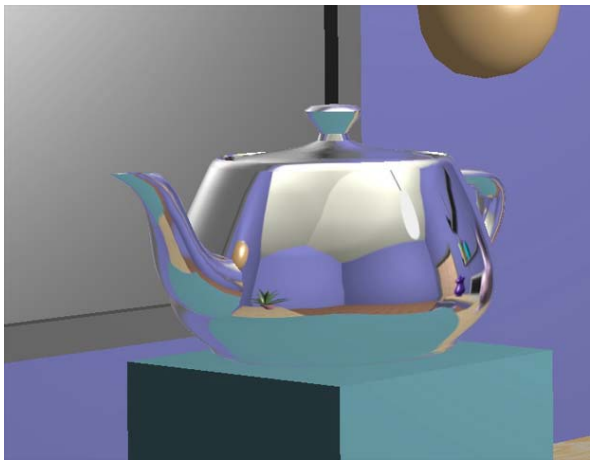


Figure 43: An Exemplification of the Algorithm 'environment mapping' Using the Notorious "Utah Teapot"

Some characterizations, then, hold for pictures because of those general structures of perceptoid signs while others are just contingent consequences of one particular and idiosyncratic instance (of many equivalently possible such instances) of a perception apparatus with sufficient complexity, or of the specific language and sign systems established for those picture users. The latter properties may change with cultural development, and may even be used to investigate cultural differences and developments (being the classical fields of history of arts, and cultural anthropology, of course). Modifications in the former attributes however do result in a different concept altogether, something that is not characterizing pictures anymore but something else.

The difference between the two types of properties is clearly expressed when corresponding forms of explanations are juxtaposed: "Since visual perception follows this or that rule (for us middle-aged Europeans at the beginning of the third millennium)" – think of perspective, for example – "that and such is true for pictures (for us)" *vs.* "Since perception (in general) can only be rationally conceived if this and that relation holds" – e.g., between sender, receiver, sign vehicle, and sign content – "perceptoid signs are only possible if that and such is granted." Computational visualists are expected to know something about the first kind of explanations, but ultimately they must know everything about the second kind by heart if they want to earn their name.

From a general act-theoretic perspective, two lines of argumentation have been followed, associated with levels of understanding with growing complexities: for (*i*) perception and (*ii*) sign use. Combinations have been sketched on different levels, but only the most complex pair – perception of individual things and assertive language – allows us to reach the conceptual structure of perceptoid signs. The goal in the next chapter is to take the essential structures resulting for pictures as a guidance toward the specification of the "complete data type »image«" that underlies and structures every effort in computational visualistics, if one such thing can rationally be considered; or alternatively to motivate the set of distinct data structures necessary, and to clarify the relations they stand in.

The gigantic task of unfolding the new discipline on the basis of general characterizations of visual perceptoid signs can indeed act as a research programme only in the present context. As has been mentioned in Chapter 2, computational visualistics, as a coherent field of computer science, is a relatively new idea, a consequence of general visualistics – as a unified image science – not being established earlier. Up to now, dealing with pictures in computer science has been separated in several sub-disciplines with more or less loose methodological connections (mainly by means of computer science

but not by the common subject "picture"). Only aspects of the "complete data type" (or the structured set of several types) have been considered.

Let us conclude the condensed review of theory in image science by summarizing the most essential points for the computational visualist.

- *An explication of pictorial communication must consider the double nature of perceptoid signs: their general symbolic aspect and their particular perceptual aspect. The way pictures are interpreted is mainly influenced by the way these two aspects interfere: pictures stand in the diverging force fields between communication and immersion.*
- *When dealing with pictures, one has to answer in general the question "who is communicating with whom"; that relation between sender, sign, and receiver is furthermore complicated by the very nature of conscious communication – to internally anticipate the interlocutors. Generic models of senders and receivers anticipated by the actual communication participants regularly interfere in the sign process.*
- *Pictures are signs with representational function. However the naïve conception of a simple relation "similarity"* per se *between picture and objects or states of affairs as the basis for that function does not satisfy. Internalization of the resemblance relation leads to the integration of the complicated principles of primary and secondary object constitution into the act of signification.*
- *Pictures are neither true nor false; they are used in sign acts that are authentic, i.e., legitimate, or not. This is a relation not merely between a picture and what it is used to stand for (objects, states of affairs), but a relation that additionally (and more importantly) includes the participants of the sign act (sender and receiver).*
- *Pictures are basically used as context builders: their presentation is an unsaturated form of communication that refers to complementing sign acts. The complements are not part of the picture, and can therefore not be predicted by the picture vehicle alone.*
- *Pictures are not primarily used for referring to objects, exemplifying abstract entities, or communicating states of affairs: they are employed to open up a medial discourse universe with objects as partaking in states of affairs. Reference to objects is performed by means of nominations (that depend on contexts introduced priorily). States of affairs are communicated by assertions.*
- *Objects (referred to by means of a nomination) may be used in a metonymy as exemplification for an abstract aspect. The metonymy can be extended to a corresponding picture leading to abstract pictures. Fields of concepts, which structure state of affairs, can be mapped on each other by metaphorical transfer. The metaphor can also base the use of pictures leading to structural pictures. Finally, quoted pictures and pictures of art are employed in a very special mode of usage different from the usual un-reflected mode.*

4 The Generic Data Type »Image«: General Aspects

> *By analogy, we propose that pictures be thought of as instances of* [several] *Abstract Data Types on which certain operations are defined, while other operations are not defined. For example, for many maps the operation* measure distance *is not defined. If a viewer measures a distance on the map and computes the distance between the points in reality on the basis of the scale, a wrong answer will undoubtedly result.*
>
> [STROTHOTTE *1998, 404*]

We have to deal with the data type »image« and the data structures including that type or being related to it. As was explained in section 2.2, an abstract data type is the formal equivalent of a concept in the context of data processing. Many of the details sketched in the previous chapter appear here as properties and relations within a data structure containing a version of the type »image«.

There is by no means any necessity of assuming exactly one unique date type »image« for every task in computational visualistics, or just one corresponding data structure. Although covered by single expressions with very similar ranges of uses in many cultures, the phenomenon of pictures and the contexts of use they partake are quite diverse, as was indicated in the first section of Chapter 3. In the last section of the same chapter, the differentiation between three quite general types of pictures used by SACHS-HOMBACH [*2002, Sect. 7*] has been mentioned, which could be taken as a hope that we may also be at least able to concentrate on only three main data structures with different but still similar versions of their central data type.

It is characteristical for the theory of data structures to use “inheritance relations” along lines of abstractions: very general data structures with only a few very basic rules restricting the properties of the types contained and the operations that relate them with each other can be used as a common structure underlying several different data structures with more complicated specifications. Most computer scientists are familiar with the application of such “generic data types” in object-oriented programming languages. For example, a generic data type »number« is often defined with an unspecific relation “addition” only – there are few general rules restricting how the result of an addition actually relates with the items added: the operation has to be closed, symmetric, transitive. This framework is differently filled for specific types of numbers.

Correspondingly, distinct levels of “genericity” for several generic data types »image« and the corresponding data structures may be assumed. The most abstract one (i.e., “THE” generic data type »image«) has but a few attributes, relations, and operations with a small number of restricting rules. But these most elementary structures are inherited to sub-types, as for example »representational pictures« or »reflective pictures«, with further specifications or modifications.

4.1 The Organizational Principle of the Discussion

This chaper is structured by aspects derived from the semiotic background of the generic data type in question. The course of discussion follows the conceptual triple of *pragmatics*, *semantics*, and *syntax*, which is closely related to the superimposed concept »sign« in general. They bring into the focus of interest domains of questions with more or less restricted horizons. It is an educated guess that the effects of the specific

difference ‘perceptoid’ of pictorial signs can be demonstrated particularly well by means of these domains.

As the most complete one of the three, a pragmatic investigation deals with the complex formed by a communication act and the other related acts (of communication or of any other sort), i.e., the embedding of the sign act in the “living practice” of the sign users. Which other act or behavior can or must precede a certain sign act? Which others may or have to follow? That, for example, the utterance of an assertion can be answered either immediately by an accepting or refusing (doubting) act of communication of the interlocutor, or by first performing a referential anchoring with corresponding sensory-motor test routines, this is the structure of that particular “language game”, and hence a theme of pragmatics.

More restrictive in its perspective, semantics focuses on the relations between sign vehicle and what is represented by means of it – as far as those relations are relevant for the sign use and can be investigated essentially without looking at the pragmatics of the sign act. Expressive or appellative aspects of sign uses are not taken into account. Usually, such relations are subsumed under the expression “meaning”. The links between a definite description (“this blurred photo”) and the object referred to, between a certain part of the distribution of color on a screen and the face depicted, or between a particular arm movement and a child’s emotional state are typical examples for the “objects” semantics is interested in. From a strictly semanticist point of view, indexical signs offer causality, and iconic signs similarity as candidates for a non-pragmatic meaning relation. Following the linguistic turn, it is now widely accepted that the concept of a strictly semantic investigation of meaning is ill-formed (with the exception of formal and artificial languages), and that semantics forms a special part of pragmatics focusing on the representational aspects of sign acts.

The syntactic domain of questions has the most restricted horizon as it deals, strictly speaking, only with the relations between the sign vehicles of a sign system. Since sign vehicles are essentially some physical objects used in a particular manner, syntactic investigations examine in consequence just the relations between and properties of physical objects. Representational, expressive or appellative aspects are ignored. The main focus of interest lies in determining the range of deviation of physical properties not changing the identity of a sign, and the rules of composition forming vehicles for complex signs from vehicles for simple signs. Obviously, the criteria used to distinguish one sign from the other can only be derived from the sign system as a whole, i.e., from a pragmatic point of view.

Although the „natural“ order is, thus, to start with pragmatics, and then progress to the particular aspects of semantics and syntax, we shall go the other direction – a procedure quite familiar to computer scientists, as programming languages are usually explained by starting with the syntax, adding the semantics of the constructs, and sometimes complementing the explanation by style guides as a very weak form of pragmatics.

In our context, the section on pragmatics (4.4) serves the purpose of bringing the participants of the pictorial sign act, their interest in those images, the purpose of their interactions with the computer, and the general communicative setting into prominent focus. Before doing so, section 4.2 investigates the options to reproduce pictures by means of a computer: (a) How can this particularly complicated but nevertheless closely restricted physical artifact provide the range of properties that renders it useful as a picture vehicle? (b) How does that effect the data type »image« that ultimately determines the

subject of computational visualistics? These syntactic considerations are followed by focusing on the representational aspects of pictorial communication: under the label "semantics" we investigate in section 4.3 a particular set of relations between the data type »image« and other data types covering the image's content.

4.2 Syntactic Aspects

Combining elementary sign vehicles into complex ones is often viewed as the central issue of syntactic investigations. In this tradition, STROTHOTTE & STROTHOTTE [*1997, Sect. 3.1*] have presented some thoughts about a combinatorial syntax for computational pictures. They have introduced analogies that may be drawn between linguistic and pictorial levels of sign elements we shall also use in this section. In particular, we have to distinguish between the non-autonomous elements combined into a picture (4.2.2.1), and the combination of autonomous pictures into pictorial (or other) signs of a higher order (sect. 4.2.2.2).

In the present discussion about the concept of pictures, the property of having a dense range of sign elements is considered prominent among the syntactic aspects: verbal signs are characterized in contrast by their discreet succession of elements. GOODMAN has introduced in this context the concept »density« – intuitively related to the structure of rational numbers – by means of a strange bipartite negation concerning (*i*) the uniqueness of the relation between sign vehicle and sign, and (*ii*) the existence of an effective procedure to prove the former relation. However, the two negations also hold for the concept »continuity«. As the distinction between the two concepts throws some light on the other two themes of pictorial syntax in computational visualistics, it is discussed first.[31]

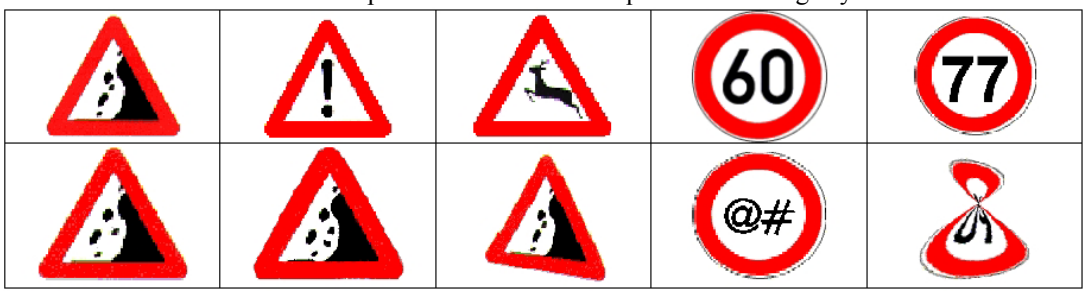
4.2.1 Pictorial Resolution and the Identity of Images

For GOODMAN [*1976, 133*] it is essentially the attribute of *syntactic density* that is characteristic for pictorial sign systems, and hence plays an important role for the corresponding data structures. A sign system is called syntactically dense, if the dimension of values for at least one of the syntactically relevant properties of the sign vehicles corresponds to the rational numbers: between any two values there are always more values. Sign vehicles with different values in that property are taken as different signs in that sign system. That is, two of the infinitely many signs of such a system can be "infinitely similar" to each other. If no such dimension of properties is given, that is, if all syntactically relevant attributes take values that can be separated from each other distinctly, the sign system is called *syntactically discreet*.

Syntactic characteristics of pictures are obviously defined by the visual properties of a marked surface of the picture vehicle. There are at least two different relevant dimensions that are apparently dense: (*i*) the positions of a point of color or a border between colors, and (*ii*) the perceived color (in a broad meaning). In the following, the range of positions and its connection to the concept »resolution« is investigated (for color cf. sect. 4.2.3).

[31] Let us, for the time being, restrict ourselves to flat, smooth, rectangular pictures. The other forms can usually be dealt with in analogy or by a simple projection to the basic form.

Table 1: Five picture vehicles of two quite different sign systems



If syntactic density is accepted as an important criterion for pictorial sign systems, the author has not given pictures in the first row of Table 1 at all. The second row, however, is used in fact for exemplifying five pictures.[32]

The signs in the upper row belong to the discreet sign system of the international traffic signs. The first three sign vehicles depicted in the lower line can indeed be also used as vehicles for the first sign in the upper row. However, those three sign vehicles carry three quite distinct signs when viewed as pictures – the sign system actually intended in the second row. The syntactically discreet system of traffic signs is embedded in the syntactically dense system of pictures: each traffic sign forms the island of an equivalence class, so to speak, surrounded by pictures that are not used as traffic signs. Similarly, the sign vehicles of letters – "a", "b", "c", etc. – can be conceived of as pictures (as in a font editor) and also as signs in a syntactically distinct sign system (as in a word processor) depending on the pragmatic context.

The syntactically characteristic property of density is of high significance for the possibility of encoding, presenting, storing, and transferring pictures in/with a computer. Is it decidable whether two pictures are syntactically equal? Can we, with other words, determine whether the transmission of a picture through the Internet, for example, has been correct, or whether a stored image still corresponds to the original?

GOODMAN deduces from density as the relevant syntactic property of pictorial sign systems that sign vehicles cannot be associated uniquely to one sign alone. In consequence, an effective proof of correctness for any image transmission becomes impossible on the base of syntax. He writes, referring to signs ("characters") of a sign system that are determined by the lengths of sign vehicles ("marks") [*1976, 132*]:

> *Corresponding to the different rational numbers, there will, then, always be two (or more precisely: infinitely many) characters such that measuring cannot determine that the mark does not belong to them independent of how precise the length of a mark can be measured.*

Here GOODMAN implicates a division of problems in classes of decidability well-known to any computer scientist. In the concrete example, a *semi-decidable* problem is considered. We have to decide whether a length is not equal to another length.[33] The positive case ("not equal") can be determined with a finite number of steps: by comparing with successively higher resolutions. That does not work for the negative case of the

[32] The reference to the sender is of course quite essential here, as a finite set of example sign vehicles alone does not properly identify a sign system – we would at least have to add "that's all". The property of syntactic density or discreetness can be ascribed only to the intended sign system as a whole.

[33] The same holds true if not lengths but the positions of spots of color are considered.

question (not "not equal", i.e., "equal" – here is indeed the reason for the complicated formulation with double negation GOODMAN uses).

4.2.1.1 Density, Continuity, and Decidability

From a formal perspective, density can be stated if there is a property of the sign vehicles (like position) that is structured as to allow us to speak about a "between"-relation of its values. This "between"-relation must again fulfill certain conditions. That is, syntactic density is a property of another property ("between") of attributes ("position") of objects. Rational numbers are the archetype for density. Real numbers are dense as well, but they have also another property of the same general structure: they are continuous. A type of numbers is continuous if the *limes* of any infinite sequence of numbers also belongs to that type. This is not true for the rational numbers, as for example the sequence of numbers approximating the relation of a circle's diameter and its circumference has as its limit not a rational number. The real numbers can indeed be conceived of as introduced by means of closing the rational numbers under the *limes* operation.

Being material objects, picture vehicles have surfaces we usually consider as being continuous, i.e., associated with the real numbers. Marks on that surface, spots of pigment, for example, may be the results of movements (of a brush, a droplet of ink, a jet of electrons). In physics, we have to consider a continuous range of locations in order to describe the interception of the movement with the surface adequately, i.e., without paradoxes of ZENO's type haunting our conception.

The distinction between density and continuity for the range of positions of pictorial surfaces is particularly important because the two types of numbers are associated with different kinds of infinity. The rational numbers can be enumerated while the real numbers cannot [CANTOR *1874*]. As is well-known, many problems concerning the question whether an instance with a specific combination of attributes does exist can be decided if the members of the set considered can be enumerated: any member can, then, be reached for checking after a limited number of steps. Correspondingly, testing the identity of any two numbers (e.g., in decimal notation) is a decidable problem only for the rational numbers. For real numbers, the test is only semi-decidable: we can find out in a finite number of steps whether two numbers (of usually infinitely many figures) *are not* the same, but in general not whether they *are indeed* the same.

The observation that picture vehicles must be viewed as a field of concept with a continuous, hence over-enumerablely infinite range of locations due to the conditions of their production does by no means imply the type of infinity for the locations relevant for the concept »image«. Although the vehicles may be linked with locations by real numbers, it is still possible to assume that rational numbers are sufficient for the range of positions relevant for pictures: densely ordered equivalence classes in the continuous sea of possible picture vehicles.

4.2.1.2 Syntactic Types of Pictures in Computers

Based on the classes of decidability, three classes of pictures can be distinguished on syntactic grounds in computer science.

A) Simple pixel pictures – the data type »bitmap«

The most simple and well-known type for making pictures available for a digital computer are bitmaps – matrices of *pixels* as they are called ('picture elements'). This data type allows us to define a pixel-value for any pair of coordinates taken from two finite sets of successive indices (i.e., natural numbers). The pixel values encode a visual property, like color or intensity (Fig. 44). Bitmaps have therefore a finite and fixed locale resolution that depends on the size a pixel is given: bitmap pictures are ratcheted. The presentation of pictures on a computer screen usually employs this data type (essentially in just one matrix size).

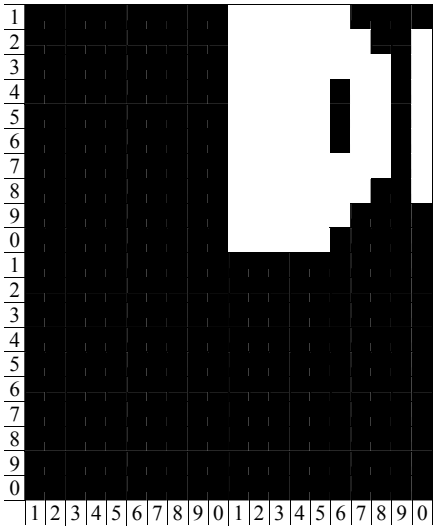


Figure 44: Schematic Bitmap

The number of different bitmaps of a given index size is finite, while the number of different index sizes is infinite but enumerable. Although bitmaps are a rather limited candidate for the data type »image«, they have at least the advantage that there is no problem to decide identity or difference between two instances effectively.

B) Pixel pictures with variable resolution

Other candidates have been developed that do not use a fixed resolution – originally in order to save memory space (and to minimize transmission times). Starting with a simple bitmap, the granularity can be reduced to just one pixel for a connected region with the same pixel values. If the description of the region is not too complicated, this results in a dramatic reduction of memory space or transmission time. Dividing a matrix recursively in halves (or quarters) provides a good algorithm to find out promising regions without a complicated specification: the description of the regions takes the form of a binary (or quarterny) tree. Figure 44 gives us a quite extreme example: this 20 * 20 matrix with its 400 pixels can be reduced to: one value for the left half, one value for the lower half of the right half, one value for the left half of the upper-right quarter, etc. If each remaining bitmap is quartered instead of halved so that we do not have to bother about alternating the direction of the parting), the resulting data-type is called a »quadtree« (cf., e.g., [FOLEY *ET AL. 1996, 843 ff*]).

Of course, the original idea of locally reducing the granularity of a given bitmap for the internal representation of that image can be inverted: the recursive definition of »quadtree« and similar data types allows us to increase the resolution of a simple bitmap at relevant locations if necessary – up to the ultimate degree of resolution for any sub region (Fig. 45).

Let us call all instances of »quadtree« with the same frame size a "quadtree family". Each element of a family has obviously a finite maximal resolution; but there are always members with a higher resolution. This property of »quadtree« should remind us of the rational numbers. The data type indeed determines a syntactically dense domain: between any two different quadtree instances of the same family there are always other instances (e.g., with higher maximal granularity), indeed infinitely many others, that is. Nevertheless the finite maximal resolution of each instance opens the possibility to

check identity and difference of two quadtrees effectively. Since there is no resolution fixedly associated with the data type »quadtree« (and its relatives), all members of a quadtree family can easily be compared effectively with each other. With quadtrees, we are able to decide whether two instances are indeed versions of the same bitmap at different compression rates, so to speak.

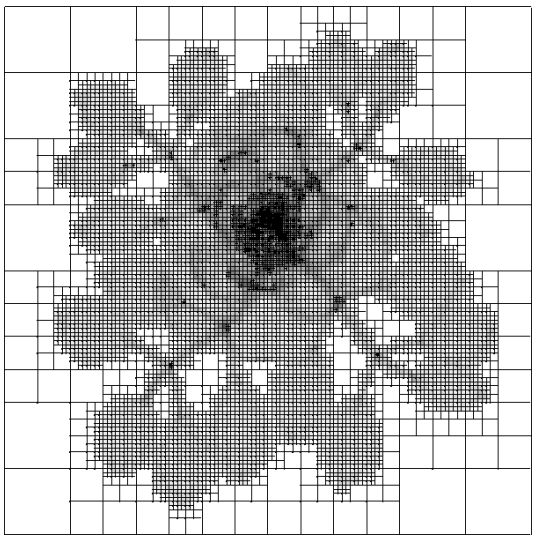


Figure 45: Quadtree Structure with Nine Levels of Detail for the Picture of a Rose

In contrast to simple bitmaps, the data type »quadtree« can be used to grasp infinitely many different pictures of a given frame size. The relation between the abstract incarnation of pictures by means of »quadtree« and bodily pictures corresponds to the relation between rational numbers and rational lengths: in the abstract structures of the descriptive systems (»quadtree« or »rational number«), two instances can be clearly distinguished despite the density; for rational numbers embodied by physical lengths and for bodily pictures alike, the decision problem is only semi-decidable.

In order to view quadtree pictures it is still necessary to transform them to simple bitmaps corresponding to the uniform and finite resolution of the video screen (or printer). The variable resolution can only be made available by means of an explicit zooming operation, which projects a local refinement of granularity onto the bitmap level. We shall come back to the zooming operation soon.

C) Generative computer pictures

Real numbers can be adequately represented by digital computers only in an intensional manner, i.e., by means of rules for sequences of other (rational or natural) numbers. There is, so to speak, no purely syntactic version of computerized real numbers. This of course reduces the chances for continuous computer pictures dramatically, leaving open, however, the option of a generative procedure that generates *always* on request a new "surface picture" with a refined resolution. Imagine a quadtree with infinitely many branching levels but without leaves.[34] It is, then, reasonable to speak of a "real" image with infinite resolution underlying any "observable" incarnation. Such generative data types are indeed of the same infinity type as the real numbers: they form a syntactically continuous domain of computer pictures. Some of the programs for generating the popular fractal pictures (Fig. 46) belong here. In principle (i.e., from a structural point of view, ignoring the familiar technical restrictions), the users can zoom into such a picture at any place without limits – on the screen, they always get a provisory version of the actual continuous picture. The generation rules are pretty simple for fractal images. Pictures more useful for everyday life would need numerous much more complex rules (which also have to include semantic and pragmatic factors).

Those are already all the relevant syntactic classes of representing pictures in a computer. If we want to ascribe to pictures the syntactic attribute of being continuous, we have to count with severe restrictions since not all the pictures are representable with a

[34] In fact, the definition of »quadtree« includes that the number of branching levels is limited.

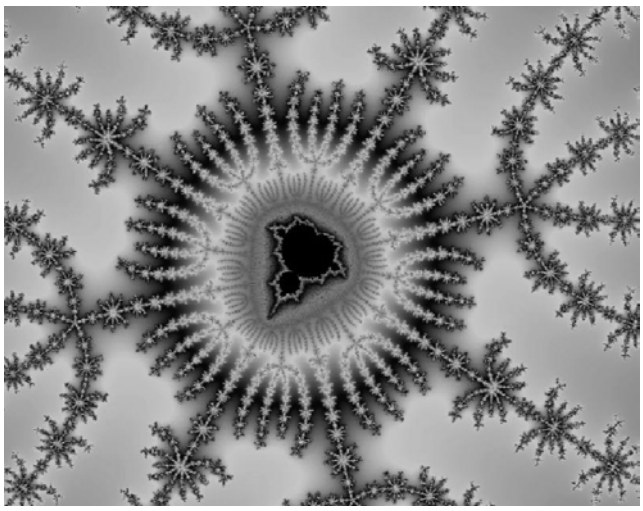


Figure 46: Simple Fractal Picture („Mandelbrodt set“)

computer. Furthermore, we do not even know whether or not a refined resolution suffices to tell us about a difference since the differences between two pictures may be infinitesimally small.

We could confine our concept »image« to syntactic density. Then, the data type »quadtree« suffices us to deal computationally with all possible pictures (presuming that enough resources are available). We can also develop algorithms that decide, for example, whether a transmission has altered a picture.

Are there other reasons to consider a dense range of locations for the data type »image«? Or is it nevertheless necessary to aim for a “continuous syntax”? Indeed, not what pictures *per se* are syntactically is the question, but what do we want to do with them, what kind of “room for action” do we want to use, and how do we want to communicate about that. For a satisfying conclusion between the arguments supporting pictorial continuity, and others speaking for a restriction to density, the relation between pictures and perceived (and depicted) reality is to be considered. Indeed, as was marked above, a certain behavior associated with pictures and with visual perception in general – i.e., an aspect of pragmatics – is of particular importance for the question of syntactic classes: the *zooming operation*, which is a generalized version of any beholder’s option of moving his/her visual sensors closer to or away from the scene observed.

Considering a zooming operation has the particular advantage in computational complexity that it is at each moment necessary to deal only with a finite number of locations – as there is analogously a limited number of sensory cells only. The locations “in between” can be “generated” if necessary by means of the zooming operation – performed either by concretely approaching the scene, by using an optical device (telescope/microscope), or by performing an algorithm controlling the computer screen. The unusual inversion of verification of ascribing positions (cf. Sect. 3.4.4) is also closely related to the ability of moving and orientating the optical sensors. With other words: the field of concept for visual perception, which is connected so closely with pictures by their definition, depends on the concept »motion«, which can only be described in a continuous domain.

The general dependency of syntax from pragmatics thus gains a particular meaning when considering pictures: density or continuity as a syntactic attribute of the underlying faculty of visual perception can only be determined as depending from a certain type of behavior – the ability to move the visual sensors in space.

4.2.2 Remarks on Compositionality

The infinity class of the parameter “resolution” is only one aspect of pictorial syntax. It corresponds roughly to the level of linguistics dealing merely with the range of letters; the notorious pixel usually comes into the beholder’s (or creator’s) focus of attention only when the presentation quality of a picture is low. There are other parts of which a

picture is viewed as composed of and which could be rearranged to form another image – thus forming the basis of a morphology of pictures, so to speak. Furthermore, several images can be arranged into pictorial signs of higher order, mimicking the arrangement of words into sentences and texts.

4.2.2.1 Composition of One Picture: Pictorial "Morphology"

The linguistic branch of morphology investigates essentially how words are build from "morphemes" – minimal meaning-contributing particles, like the postfix '-ed' in English, the prefix 'pré-' in French, or the stem '-wend-' in German. Mostly, such morphological elements are identified and arranged into classes by means of a rule of interchange: some words beginning with 'pré-' can be transformed into other words of French by just changing the prefix to 're-', 'con-', 'de-' etc. The morphemes may best be viewed as the vehicles of unsaturated partial signs acts without a pragmatic function of their own (unlike predication or nomination) that modify in a more or less specific way the meaning of the whole.

Are analogous "pixemes" relevant for the generic data type »image« or any of its more specific derivatives? It is important to note here that semantic arguments may be used to find such pixemes, but that their description must avoid any semantic "contamination". The characterization of pictures as *perceptoid* signs of the visual sense modalities already suggests that visual Gestalt entities may serve exactly that purpose: closed areas, grouped by neighborhood and similarity (e.g., of coloration); connected lines; some visual pattern inducing directional "energy" (diagonals, arrow shapes).[35] On a more formalized level, we may consider geometric entities – lines, curves, dots, areas, etc. as the basic morphological components of pictures. Indeed, such entities are also the standard elements offered by painter programs (like *Corel Draw*).

Let us concentrate for the moment on lines or strokes. A stroke may be defined pragmatically by the painter's movement or semantically as the contour line of an object. Beside the potential graphical meaning of a line or the stylistic indications associated with its particular make (not to mention any other expressive or appellative function of dynamism associated to it on the level of pragmatics), there are several dimensions in which a line – just being taken as a line – can vary: most prominently in the course or path it takes. But there are other ranges: is it a continuous line, or dashed, or dotted? Does it consist of strokes of one kind or another? How thick is it? Does its thickness change over its course or not? Is there an internal fine structure to the strokes? Assuming a corresponding data type »pictorial line« separate from »image« is, thus, certainly a wise idea.

An extensive treatment of such a data type and its possible implementations has been performed in the context of non-photorealistic rendering (NPR) , a sub field of computer graphics. While Figure 47 exemplifies several types of digital "hairy brush strokes" that have been generated – quite expensively in computational resources – by simulating a brush with several individual bristles applied with changing pressure, Figure 48 shows examples of lines resulting the application of a "style function" to the "skeletal path" of the stroke. Both constituents of the latter case are defined by means of parametric curves: the style describes how a given path (as the core of the line) is to be perturbed in order to result in a corresponding pixeme. Style and path can be viewed as independent ranges determined in each particular picture by semantic and / or pragmatic aspects.

[35] [SACHS-HOMBACH *2002, II.4.3*] offers an overview on important "classical" texts for that theme.

Figure 47: Enlarged Fine Structure of Computer-Generated Stroke Types

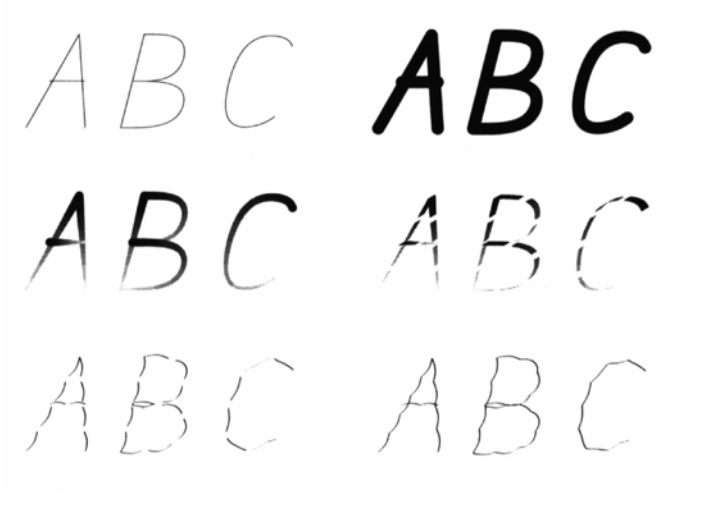


Figure 48: Examples with Style-Parameterized Stroke Functions

The rules of composition of strokes or other pixemes into a picture can be investigated by means of the tools of formal languages. Every computer scientist knows by heart the structures called formal grammars – or CHOMSKY grammars – since those are the major instrument for defining and classifying linear structures like programming languages. Formal grammars based on replacement rules that lead to two-dimensional "pictorial" structures have been investigated essentially under the name of *L-systems*.[36] The expressions generated by an L-system can be interpreted as orders to place substructures, and to move or turn in-between. A fairly simple example is defined by the following replacement rule:

$$P \rightarrow P\,[\,-P\,]\,P\,[\,+P\,]\,P$$

Interpret "P" as "place a pixeme and move a bit forward", "+" by "turn right", "–" by "turn left", and the square brackets as stack operations that allow us to return to that point after the bracketed sub expression has been dealt with. The plant-like structures in Figure 49 have been generated by this rule. Obviously the pixemes themselves are not really relevant for L-systems and their relatives, since these grammars basically deal with arrangements and groupings of abstract entities that may or may not be interpreted in a pictorial sense.

For a more extensive approach to pictorial morphology, a data type for pixemes can best be derived from a calculus for geometry. That any pixeme must be a geometric entity seems almost too trivial to be mentioned. That inversely any entity in flat geometry – apart from non-extended points – may also be a candidate for a pixeme is at least a good guess. Taking the common Euclidean formalization of geometry leads however to the "unpleasant" consequence that the most basic pixemes must be non-extended points – a concept highly abstracted from experience, that is. Non-standard approaches to geometry like mereogeometries[37] here offer an interesting way out. The traditional calculus

[36] The "L" stands for "Lindenmayer", as the botanist ARISTID LINDENMAYER started to use a corresponding formal language for describing plants; cf. [PRUSINKIEWICZ & LINDENMAYER *1990*].

[37] cf. [WHITEHEAD *1929*], [LEONARD & GOODMAN *1940*] [CLARKE *1981*], [AURNAGUE & VIEU *1993*], [ASHER & VIEU *1995*], [SMITH *1996*], and [BORGO & MASOLO *2001*].

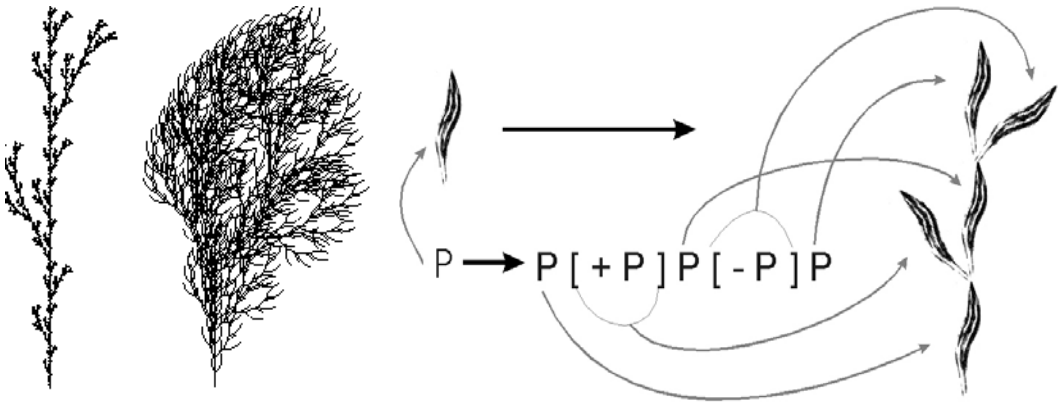


Figure 49: Two Example Pictures Generated by (Bracketed) L-Systems, and the Graphical Interpretation for the Rule for the Left Example

of geometry develops around the fundamental concept of a zero-dimensional point. In contrast, mereogeometries are based on extended regions as the most elementary entities, which may or may not have (distinguishable) proper parts. The regions are often called "individuals". Individuals do not have immediate attributes of form or position: only the relations to other individuals, in particular parts, determine form and (relative) location.

An individual may quite well be thought of as a visual Gestalt – thus following the principle of perception psychology of the Gestalt school: one has to consider the perceived whole first and introduce the concepts for perceptual atoms as instruments of the explanations of the former, not the other way round. We do not see sets of zero-dimensional points but regional Gestalts. The abstract notion of a spatial entity without extension is secondarily constructed in order to explain some aspects of experienced space, but leads on the other side to severe difficulties as the discussion on infinite resolution has shown. The thesis is therefore that the constructs of an individual calculus for the two-dimensional mereogeometry are excellent candidates for a general and exhaustive discussion of pixemes.

The syllable "mereo" indicates that part-whole relations form a central aspect of mereogeometries: more precisely, the fundamental data type "individual" in mereogeometries is primarily characterized by the reflexive and transitive relation of being part of between two of its instances.[38] In the words of B. SMITH [*1996, 290*]: "*We adopt as mereological primitive the relation of parthood or constituency. We say x is a part of y, and write 'P(x, y)', when x is any sort of part of y, including an improper part (so P(x, y) will be consistent with x's being identical to y).*" With this relation, more complex relations and entities can be formally defined, especially those with topological interpretations, like boundaries and interiors. Two individuals are, for example, defined to "overlap", if there exists a third individual being simultaneously part of both of them. In particular, the concept of a minimal region usually called a "point" ("Pt") – we may well use "pixel" instead – can be introduced: $Pt(x) =_{def} \forall y (P(y, x) \Rightarrow y = x)$. That is: a

[38] Some mereotopologies and mereogeometries are based on other relations; for example, AURNAGUE and VIEU [1993, 403] use the symmetric, reflexive binary relation *C* (for "being connected with"), from which (among others) the relation used in the text, P (part of/inclusion), is derived: $P(x, y) \equiv_{def} \forall z (C(z, x) \Rightarrow C(z, y))$

point in this sense is a region that has no proper parts (or rather, a region where no proper parts are considered).[39] [40]

When the concept »point« is introduced in the data structure as mentioned above, there is no need in any concrete instance for using infinitely many point instances: only the "relevant" points must be instantiated. This also means that there is always a finite resolution. N. ASHER & L. VIEU [*1995*] propose a formal mechanism called "microscopization" covering a kind of zooming operation by means of a modal extension to their calculus. What is a "point" on one level may be a compound of regions with several points on a microscopized level. While Euclidean geometry first introduces the continuous range of infinitely many coordinates determining potential points some of which are then chosen to be relevant (still an infinite number in any practical relevant instance), mereogeometry starts with a (usually finite) number of relevant individuals (regions) we can think of being given in perception. That is, we may indeed assume that the principles governing visual perception determine the regions that are syntactically relevant, hence leading only to the essential "points".

Mereogeometries are a formal way to deal with geometry in a manner more closely related to visual perception than traditional point geometry. If we accept the view that the central data type of a two-dimensional mereogeometry determines what is a pixeme – namely any connected sub system of individuals, then there is indeed no finite number of possible pixemes – a clear difference to verbal sign systems with their strictly limited number of morphemes. However, any pixeme can be described and dealt with in a unique and generatable manner in the calculus in a finite number of steps: pixemes can be combined to form pixemes of a higher order – until every visually separable Gestalt of a picture is covered.

4.2.2.2 Compositions With Pictures: Pictorial "Text Grammars"

Considering compositions of (or with) pictures to form signs of higher order brings us first back to the compositions of pixemes as by L-systems: did we not in fact arrange pictures of strokes by means of a formula derived by an L-system? Indeed, the arrangement could be, as in that case, one performed in the picture plane as well as one in our usual three-dimensional environment, or even in the separate dimension of time. While such formal systems may be also quite useful for describing part-whole relations in the sense intended here, the two forms of compositions – within one picture, and with several pictures – must not be confounded: pixemes are never used for autonomous signs, while the composition with pictures depends on the status of the component pictures of being quite well useable as independent signs. The linguistic counterpart to the latter is indeed text grammars dealing with the composition of texts from sentences, which can be seen as being already the sign vehicle for a complete sign act, a status not ascribable to single words or phrases. An abstraction of pure syntactic classes analogous

[39] There are other ways to introduce a similar notion of a point in other mereotopologies/mereogeometries, some of them leading even to the non-extended Euclidean version. The nub here is that points are logically secondary entities.

[40] So far, only the definition for a mereotopology has been sketched: by adding, for example, relations of relative distance between points (point A is closer to point B than to C), and of relative direction between points (point A is between points B and C), the data structure can be extended to a geometry basically of the same expressive power as Euclidean geometry. As an advantage over and above not chosing a highly abstracted starting point but a more perception-like entity, topological and metric aspect can then be dealt with in relative separation.

to »verb«, »noun«, »adjective« is not available, and may, in the light of the considerations of Chapter 3, never be in general.

Only for very restricted domains of use, an association between grammatical categories and pictorial compositions might be possible: TH. STROTHOTTE [*1989, Sect. 3.1*], for example, offers a syntactic schema in the context of maintenance instructions. Based on a formalized verbal description, an arrangement or sequence of pictures is to be generated. To that purpose, the noun phrases in question are schematically associated with elementary images of corresponding objects. The verbal groups considered correspond roughly to temporal arrangements of the pictorial compositions linked to the noun phrases that are bound together by those verbal groups. This includes the appearance of a "user's hand" for imperative moods, or of "think bubbles" for subjunctive moods. Although it is quite functional for its definite purpose, the rather small fraction of syntactic categories used indicates the limitation of such an approach. What about adjectives, adverbs or conjunctions, for example?

Of course, texts form just one-dimensional compositions: a comparable composite sign with pictures is given by (simple) comic strips, and also by films. While the former is clearly organized in several individual pictures by the "guts" between them, moving pictures do not offer a similar distinction of autonomous pictorial entities as easily. However, taking cuts or dissolves between (continuous) scenes as the temporal equivalent of inter-panel space in comics leaves us with exactly those scenes as pictorially autonomous signs, a solution not too implausible indeed (cf. [SCHWAN *2001*]).

Comics do not only come in a linear fashion; the more advanced specimens use quite complicated forms of layout, taking into account not only the two-dimensional area of possible placements of one page, but the options of using either two opposing pages with a marked jump of view, or the even further separation of a page to be turned. This is indeed not much different from general layouting, which mostly deals with texts and possibly some pictures or other elements in between – pure text layout and comics layout form just the two extremes. A type of pictorial composition in 2D, which is particularly interesting here, is given when pictures are shown within another picture, i.e., not just as a morpheme (like the stroke pixeme or even a texture map) but as an autonomous picture on top of the other picture plane. A typical example is the use of an enlargement inset framed by means of a pictorial magnifying glass.

There are compositions of pictures into high-order signs even in 3D space: think of an exhibition. The arrangement of the exhibits intends to establish correspondences and to allow the visitors to see more than just an unconnected set of pictures. For a computational visualist, a comparable task may come into view when dealing with special VR presentation hardware like CAVEs forcing him or her to coordinate the placement of pictures in three dimensions.

We shall not go here into further detail of this particular aspect associated with the data type »image«. Computational approaches to text/discourse grammars become rather helpful in later sections when taking into account more than just syntactic considerations of layouting.

4.2.3 Some Notes on Formalizing Color

Denn daß man sich etwas „Grauglühendes“ nicht denken kann,
gehört nicht in die Physik oder Psychologie der Farbe.

[WITTGENSTEIN *1984, I.40*]

The formal structure of the locational organization of pixemes is only one aspect of pictorial syntax, and in fact one not perceivable as such. Like the temporal base structure of music that can only be perceived as organizing a sequence of distinct auditory markers – difference of pitch or harmonic progression, change of volume or variation of timbre – the perception of the spatial base structure of pictures depends on visible differences: visual markers usually subsumed under the expression “color”. Indeed, color in this general sense includes hue, saturation and intensity as well as texture or even homogenous temporal variations thereof. It is exactly the change of any one of those values that induces the border of a pixeme.

The various systems to cover color (in the closer sense) formally in computer science are well-known – “color models” – and do not need a detailed description here: every painting program or system for picture manipulation offers at least RGB, HSB or CMYK. Here again, we meet the problem of formalizing a seemingly dense dimension – between any two colors there appear to be more colors. And again, we depend on a perceptual system with a limited resolution in color distinction.[41] In contrast to locative resolution however, there is no such thing in “color space” as a natural “zooming operation”: the members of some pairs of colors are only distinguishable by means of a complicated technical device like spectral analysis that has no equivalent in non-technical human behavior.[42] We may therefore take color without real simplification as a syntactically discrete dimension with a resolution just below the threshold of human perception. Correspondingly, contemporary computer systems offer a data type for homogenous colors with more than 16.5 million values (together with methods to select and manipulate them easily): two immediately neighboring color values of that system are for most humans undistinguishable.[43]

Homogenous color is the central, but not the only aspect. More often, the visual markers are given as fine-grained textures that only appear as more or less homogenous if the spatial resolution is not too high. In these cases, zooming reveals that a locale distribution of homogenous colors has in fact been used (or even fields with textures on a still finer level). However beside the zooming, textures are perceived, remembered, and even imagined not as a particular spatial distribution of (homogenous) color but as another kind of visual marker values (more or less analogous to accords in music): the system of visual markers consists of two levels.

As textures can technically be reduced to fine-grained patterns of homogenous colors, the most common way to deal with them in computational visualistics is by using a sample. More ambitious analytic solutions for a corresponding data type concentrate on characteristic structural, statistical or spectral parameters [LONG *ET AL. 2000*]. Structural parameters characterize textures according to geometric relations between correspond-

[41] Are there arguments for taking color space to be even continuous? Physics at least assumes a continuous spectrum (range of wavelengths / frequencies) of electromagnetic waves implementing color, though the relevance of this conception for color perception is only quite indirect.

[42] Zooming locational resolution by microscopes or telescopes can be viewed as a technical equivalent to approaching the scene perceived, as was already indicated above

[43] Moreover, there are few technical devices that really reproduce each single value distinctly.

ing homogenous sub regions while statistical texture parameters measure the locale variations of visual qualities (e.g., granularity, regularity, line-likeness): the feature "roughness", for example, depends on the fractal dimension of the intensity variations relative to spatial displacement (cf., e.g., [WU & CHEN *1992*]). For spectral approaches, the Fourier transform of the texture is calculated as the basis of further analyses.

For the computational visualist's perspective, transparency and reflectivity are phenomena of color (in the broad sense) even more interesting then textures. Stained church glasses or Mexican folk art with build-in pieces of mirrors are well-known examples of corresponding traditional pictures. Note that those effects cannot be ascribed to the picture as such: it has to be considered in (and in contrast to) changing situational contexts. In every single context (i.e., arrangement of objects and lights around the image), the transparent and reflective regions of the picture have a fixed appearance undistinguishable from other regions – they may be marked by homogenous colors or textures just as well. Only if changes in the context do indeed change the distribution of marker values, and hence the arrangement of pixemes, an observer perceives regions as being transparent or reflective. The phenomenon is also directly important for computational visualists when combining pictures in layout (mostly transparency) or 3D graphics (transparency and reflection). Of course, an adequate conceptualization in the data type »image« must explicitly include such "indexical marker values"; we cannot replace them by one arbitrarily induced distribution of homogenous colors or textures. As a standard for transparency, an additional dimension of marker values called the "alpha channel" has been added.[44]

Let us consider in this context as a final aspect of pictorial syntax a thesis that is often mentioned: in contrast to verbal expressions that can be syntactically ill-formed, there seems to be no such thing as a syntactically ill-formed picture (cf. [PLÜMACHER *1999*]). Whereas, for example, the syntactic structure of a verbal language may be described by just *one* Chomsky grammar, *any* expression in *any* L-system forms a picture. The reason seems to be essentially that the geometric base of pixemes is dense, and any potential combination of pixemes already forms a picture. However, those discussing this issue do usually not mention damaged screens: cuts, holes, and burned regions disrupt the homogenous topology that is part of the pictorial base structure. Cuts, for example, separate neighboring pixels: are they neighboring anymore or not, we cannot really say. Suddenly, there is non-space in picture space – which is certainly not equivalent to fully transparent regions. After all, a cut in a "Rembrandt" results not just in another picture but in a *destroyed* picture. So, our counter-thesis is that pictures might quite well be counted as syntactically ill-formed if the underlying geometric structure is disrupted.[45]

4.2.4 Syntactic Transformations and Image Processing

Computationally, transformations of the syntactic structure of pictures belong to the field of image processing – we usually interpret such a transformation as an operation from one picture to another one although the computer deals with the vehicles alone. In

[44] Reflectivity poses some particular problems, which we deal with later.

[45] As with syntactically ill-formed verbal expressions, which may nevertheless be used efficiently for communication, syntactic well-formedness is no necessary criterion for a picture to be employed: a certain art form in the middle of the 20th century, particularly exemplified by L. FONTANA, plays exactly with this deviation from well-formed images: FONTANA's "cut pictures" are reflective pictures that focus our attention on the "materiality" – or in our terminology: on the geometric base structure – of pictures exactly by means of the violation of that very basis; cf. [SACHS-HOMBACH *2002, 164f.*].

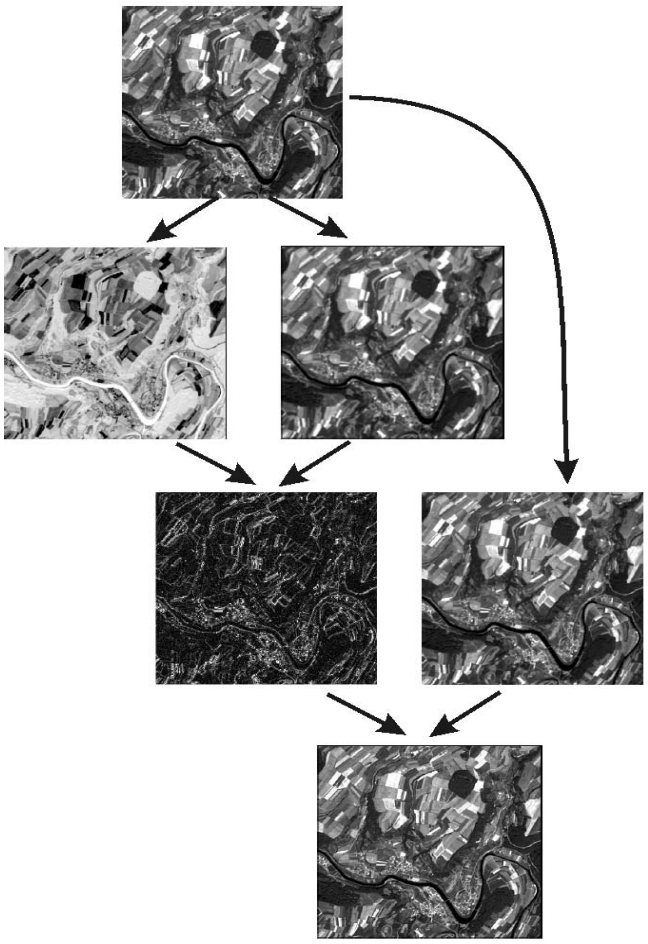


Figure 50: Sharpening Transformation in the Context of Aerial View Analysis (see text)

most cases, pixel matrixes are considered: this format is what CCD cameras or scanners originally provide.

The transformations that are performed on one image vehicle alone can be distinguished from those combining two (or more) vehicles. Typical combinations are weighted addition and difference. Apart from that, the operations can be classified into three categories: pixel operations, local operations, and global operations. In the first class, the output value at a specific pixel depends only on the input value at that same coordinate. In contrast, the output value at a specific pixel is calculated from all the input values in the vicinity of the coordinate in question for local operations – or from the complete set of input pixels for global operations.

A typical pixel operation is "inversion" leading to the negative image. Essentially, pixel operations calculate only new marker values: as in the case of inversion, the geometrical base structure is not changed. This is not necessarily so. For example, the transformation from color to grayscale (intensity) maps different marker values from one system to the same value in the second system. In consequence some syntactic elements are no longer distinguishable: the color version and the grayscale version are determined by different sets of pixemes.

Applying filters is essentially a local operation. For example, the "MEAN filter" replaces the marker value of any pixel by the average of the marker values of that pixel with its immediate neighbors (e.g., the 3 * 3 pixel neighborhood). The effect is obviously a reduction of detail – the image appears to be smoothed. For reaching the inverse effect, several operations have to be combined (cf. Fig. 50). Calculating the difference between the corresponding pixels of the original image and the image resulting a smoothing operation by the MEAN filter leads in a first step to a picture vehicle in which the marker value corresponds to zero ("black") for most pixels– the smoothed picture differs from the original only at few places. Since the mean operation has the most effect where the local fluctuation in the original picture vehicle is high, i.e., at sharp borders, the difference pixel matrix highlights essentially such edge pixels. When in a second step the original pixel matrix is added again to the difference matrix, a pic-

ture vehicle results that is like the original with the contrast heightened exactly at the pixemes borders.[46]

A special local transformation is the change of resolution, and in particular its reduction, since it often leads to unwanted artifacts: everybody is familiar with the Moiré patterns in TV that appear if a regular texture with a periodicity close to the camera resolution is shown. Aliasing effects disfigure smooth curves if they have to be digitized into a pixel matrix. Therefore, anti-aliasing transformations are usually added that smooth away the artificial borders

The most important one of the global operations is the (digital) Fourier transformation. This transformation is based on the fact that every possible spatial variation of the marker values can be described by a parameterized set of cyclic functions. Therefore, the two-dimensional spatial distribution of pictorial marker values can be transposed into a two-dimensional frequency distribution of the same marker values, i.e., a new pixel matrix. However, the pixel positions are here not interpreted as locations but as a specification for the frequency of a cyclic function, or more precisely its frequency components in the two main directions. Each pixel of the Fourier transform depends on information from every pixel of the original vehicle (and, in fact, *vice versa*).

In a way, location is spread out over the whole Fourier matrix in a similar manner as the cyclic functions disperse the frequency distributions over the whole original pixel matrix. Quite obviously, the pixel matrix of the Fourier transform of the vehicle of any ordinary picture does not bear any resemblance with the original. However, many transformations that need complicated calculations in the original picture vehicle can be calculated quite easily in its Fourier-transformed version. Therefore efficient versions for calculating Fourier transformations and their inversions, e.g., FFT, are extremely important in image processing (cf., e.g., [GONZALEZ & WOODS *2002*]).

4.2.5 The Limitations of Pictorial Syntax

In conclusion: the formal treatment of pictures in computational visualistics covering the syntactic aspects rests essentially on two basic data types and their interaction: first, the base structure of position and form, for which the calculi of mereogeometry are the most promissing general candidates; second, the field of marker values based on a discretized range of homogenous colors and an additional dimension for transparency (and perhaps reflectivity), offering further structural principles for the level of textures.

Providing structures isomorphic to the syntactic characteristics of images is indeed sufficient for *handling* pictures by means of a computer – after all that structure is exactly equivalent to all the relevant aspects of the picture vehicles. However, computational visualistist should not be satisfied, as pictures are not merely picture vehicles but much more complicated entities. Not everything flat and covered with regions of textures is already a picture. If we do not also consider the particular contexts of use that make us take a flat object for a picture, there is, for example, no way to select rationally from a given set of pictures the one to be best presented to a certain computer user under some specific conditions at hand: computational visualistics would not reach its full capacity.

Even the syntactic grouping of pixemes into entities of higher order takes into account not only the syntactic attributes of the corresponding elements but also more or

[46] A similar combined transformation is in fact encoded in the neuronal network of the retina, leading also to the optical illusion of MACH bands.

less every other pixeme present in the picture: the grouping is highly context-sensitive. Indeed, the identification of the pixemes particularly in a figurative picture depends to a high degree on the picture's content, i.e., what is depicted – a question belonging (at least on first view) to the field of semantics.

4.3 Semantic Aspects

Semantics in computational visualistics adds those features to the data structure around the type »image« that deal particularly with representational aspects of pictures. There are, in fact, two facetts of representation to be covered: by considering the "picture content" we focus on those properties of the picture vehicle that are relevant for understanding its significance in the sign act – our abilities to recognize pixemes as something more than patches of texture, so to speak. If we mention "the referent of a picture" we mean the individual scenes, events, objects, facts, etc. that the picture is taken to represent. Since we consider them as intentional objects, i.e., only as far as somebody experiences, recognizes, knows them, the referents may be factual or fictitious. They depend on the picture content. Therefore, picture content is our main focus of attention in this section. For the sake of simplicity, picture contents may be transcribed as predicative partial expressions, e.g., "being a large red suspension bridge spanning over water to a hilly countryside", while nominative partial expressions are used to mention picture referents: e.g., "the Golden Gate Bridge". That replacement by means of verbal expressions is, of course, necessary because we cannot deal here with contents or referents directly.

We can use linguistics again to inspire our attention. One rather stable distinction in linguistic semantics – though often appearing under different names – is that between *reference semantics* and *intra-lexical semantics*: researchers in the latter framework concentrate their efforts on circumscribing the representational aspects of verbal expressions in an explicit and formal manner. The content of an expression, sentence or text is then given in a meta-language.[47] A relation to extra-linguistic entities (*vulgo*: the world) is not considered in any direct form. Computational linguistics, the older sister discipline of computational visualistics, uses such translations as the internal representation of the meaning of sentences or texts in a computer. In the form of *operational semantics*, they are employed together with translation routines and transformation algorithms for simulating the understanding and generating of natural language.

Linguists dealing with reference semantics try to ground the meaning of verbal expressions in the world, in particular by investigating the role of contexts: at least those terms dealing with concrete, spatio-temporally extended and localizable affairs are usually understood as being anchored in non-verbal experiences. Assumedly, the reference relation, which associates each expression with its (usually) non-verbal "thing", is mentally mediated. Essentially, it is perception that gives access to contexts, and thus supplies the needed referents. By and large, visual perception is used as a paradigmatic case for studying reference semantics. We come back to the operational form of reference semantics in Section 4.3.2, as it determines an important part of computer vision.

As indicated above, a first approach to semantics of pictures is something analogous to intra-lexical semantics: a translation of the meaning components into a meta-language – indeed the same type of logical meta-languages used for verbal expressions. Such an approach has often been criticized: importing the categories of verbal signs for

[47] Some think here even of a so-called *language of mind*, or "Mentalese"; [FODOR *1976*].

analyzing perceptoid signs may be inadequate. However, a proper "intra-pictorial semantics" has not been proposed, not even in a sketchy form. The idea of a translation of a picture's meaning components into other pictures on a meta-level – perhaps the "pictures of the mind"? – appears not to be really promising for scientific purposes. On the other hand, transferring the conception of reference semantics directly to pictures is complicated, as well, if we take into account that images have the prime function of context building: they provide the (absent) contexts that contain the referents for verbal expressions.

Using verbal interpretations as a mediating link is a quite plausible solution in particular because computational visualists deal essentially with the concepts forming image content, not with the thing *per se*. Whatever is to be represented by pictures has to be covered in the essence in the generic data structure by some – presumably relatively unspecific – data types, which we may call »picture content« (and »picture referent« respectively). A representation relation Rep associates an instance of »image« to one (or perhaps even several) instance(s) of »picture content«[48] – a relation obviously of particular interest for computer vision algorithms and the determination of pixemes. The inverse projection relation Rep^{-1} has to be considered as a keystone of computer graphics and information visualization.

The sub-types of »image« differ in the kind of »picture content« they are related to, and the internal structure of those relations. For example, SACHS-HOMBACH's three subtypes of pictures, representational, structural, and reflective ones, are quite distinct in semantic respects. Representational pictures are used for representing realistic contexts, i.e., arrangements of spatial objects (or rather the intentional pendants thereof). The relations between content and image have to be structured in a way that realizes the characteristics 'perceptoid' for this type of signs as a more or less direct resemblance$_\beta$ – possibly modified by a metonymic shift. For structural pictures, a metaphoric shift has to be additionally considered that transforms non-spatial entities into spatial things or non-visual properties into visual ones. Finally, the meaning of reflective pictures (not too prominent in computer science, anyway) is mostly not a picture content in the close sense but the relation to other pictorial sign acts (including their semantic relations). In the following, we therefore concentrate on representational pictures.

4.3.1 Computer Graphics, Spatial Objects, and Perspective

The contents of representational pictures are essentially configurations of spatial objects. So, what *are* spatial objects, i.e., what determines the concept of material individuals that form the arrangements evoked by images? The answer was already given in Section 3.4.4: it is "sortals". The most general form of the data type »picture content« for representational pictures must cover the complicated internal structure of sortal concepts. This structure is especially important for the generation of corresponding images, for computer graphics, that is. The standart starting point for computer graphics is called a geometric model, and we have to investigate the relation between geometric models and sortal objects.

[48] In the case of reflective pictures, we have to consider even images with an immediate Rep-set that is empty – take for example M. ROTHKO's monochrome screens (but see also Sect. 4.4.5).

4.3.1.1 Sortal Objects and Geometric Models

The kind of geometric models most commonly used consists simply of sets of polygons in three-dimensional Euclidean space: the surfaces of the objects considered.[49] This form has the advantage that many of the consecutive processing steps of picture generation become relatively easy. An obvious disadvantage of polygonal models is: they do not provide a proper way for dealing with smoothly curved surfaces. However, more important at this place is the following problem: geometric models of more extensive spatial scenes with many thousands of polygons tend to become extremely hard to follow up when being edited. Although this does not seem to be a difficulty of the picture generating algorithms as such, but only one for the modeling computational visualist's efficieny of access to the data, the problem has its root in a very general simplification: the data type »geometric model« is not equivalent to sortal concepts; it is only a very coarse approximation sufficient for some aspects of computational image generation. It does not really correspond to spatial objects in the everyday sense.

For one: the "polygonal soup" as such is not internally structured. Geometric neighborhood and the sharing of common nodes are the only relations between two polygons inherent to that data type. In order to ease the editing and re-use, groupings of polygons and hierarchies of polygon groups have been added (cf. e.g., [PREIM & HOPPE *1998*]). Quite obviously, the supplementary relation is a kind of part-whole relation – the second crucial component of sortal concepts apart form geometric Gestalts, as we remember from Section 3.4.4. Correspondingly, compounds of grouped polygons are often called 'objects' already.

Even so, there is only a weak criterion of identity for groups of polygons. More precisely: whether or not two groups of polygons in separate models are the same depends on their structural organizations alone. They lack conceptually the unique spatio-temporal history of objects connecting multiple contexts: geometric models form exactly one context and are restricted to that context, similar to the pre-objects mentioned in Section 3.3.3. We cannot speak about their identity in the way expected for individual sortal objects.

The instances of the complete data structure »picture content« for representational pictures may best be circumscribed by predicative expressions, as has already been indicated above: they do not correspond to individual object instances. But they do correspond to *concepts* of individual objects. For example, a picture's content may be describable as "being a chair", not an individual chair (e.g., "the one I sat on yesterday evening"). But that concept has to include the correct individualization criterion, which geometric models usually do not provide.

This is, of course, not to say that geometric models are of no use or merely bad use in computer graphics – the impressive results speak for themselves. It is nevertheless of great importance for a computational visualist to know exactly which purposes allow for what kind of simplifications from the complete concept of »image content« given by sortal concepts. In order to better understand the relations between sortals, geometric models, and their pictorial projection, an excursion to the use of arguments between fields of concepts is necessary.

[49] Other formats in use are more or less equivalent with respect to the arguments in this section.

4.3.1.2 Excursion into the Theory of Rational Argumentation

Let us recapitulate what has been mentioned in Section 2.1 or sketched in some parts of Chapter 3 about rational argumentation so far: assertions can only be formulated with respect to abstract reference points called concepts. A concept means: a habit of distinguishing that is socially established and mutually controlled by the members of a community. To that purpose, concepts are determined ("defined") by means of formulating relations to other concepts, and thus grouped into fields the members of which determine each other. The relations between the members of a field are often called 'meaning postulates' as they do not just determine the kind of sorting, classifying or distinguishing covered, i.e., what is meant by a corresponding predication. They also express how sentences with that predication interact with sentences with other predicators. Think of the system of meaning postulates as a logical calculus. Those relations can be employed to infer conclusions from a given set of assertions: if we agree on "Socrates is a human being" and also on the meaning postulat "the concept »human being« is determined by the concept »mortality«" then the assertion "Socrates is mortal" can be inferred.[50] Such "calculations" are exactly equivalent to what computer scientists do with an abstract data structure (cf. Sect. 2.1).

Let us have at this place a quick look at one of the formalisms developed in AI for dealing with the content and referents of verbal utterances. Like every knowledge representation language, the family of KL-ONE knowledge bases consists essentially of structured sets of propositions [BRACHMAN & SCHMOLZE *1985*]. In the case of KL-ONE, the meaning postulates of a field of concepts are covered by propositions in T-BOXes, as they are called ('T' for 'terminological'). Empirical propositions are collected into A-BOXes ('A' for 'assertive'). An A-BOX also provides the referents for a new nomination. That is, an A-BOX is indeed the KL-ONE equivalent of a context. The example syllogism mentioned above then corresponds to an A-BOX that is transformed according to the assertions about »mortality« and »humanity« in the corresponding T-BOX.

Rational argumentation is any behavior that tries to settle in a community a disagreement about the validity of an assertion or meaning postulate without violence or tricks, i.e., by means of finding an agreement about the concepts to be further on used in certain contexts by the group of speakers considered (cf. [ROS *1989/90*]). First of all, the participants in a rational argumentation may compare the meaning postulates of the concepts they understand as involved in the case of dissent in question. For example, complex concepts may be analyzed into more elementary ones of that field, and the opponents find they were indeed using different definitions ("a bachelor is an unmarried man" vs. "a bachelor is an unmarried heterosexual man who is not member of a celibatarian order"). They can now decide to employ one or the other of them in future and thus settle their disagreement.

But what happens if they do not agree even on that level? This may happen when a field of concepts is too complicated to be surveyed easily, as for example the field of sortal concepts; or if it is completely new for one of the interlocutors, like the strange concepts of quantum physics at the beginning of the last century. What kind of rational argument do we have for motivating that a certain set of meaning postulates "really" es-

[50] Note that the traditionally used form "all human beings are mortal" is meant with strict necessity, i.e., as a conceptual relation between the corresponding concepts. It is therefore better to explicitly refer to the concepts instead of talking about the infinitely many instances thereof.

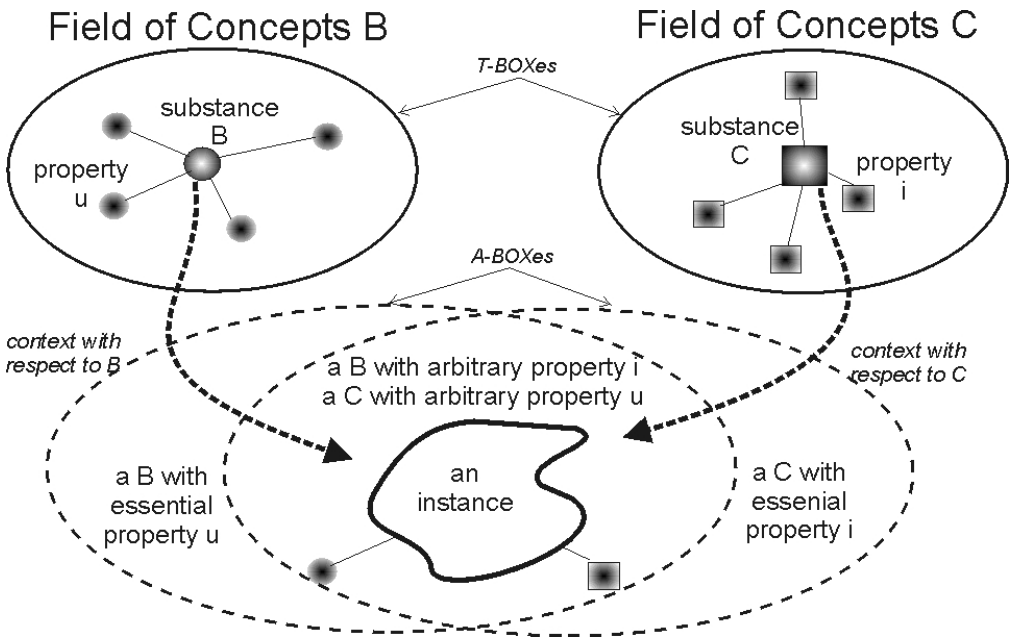


Figure 51: Graphical Schema on Ascribing Arbitrary Attributes: not a Conceptual Relation but a Phenomenon of the Transient World of Particulars

tablishes a "sound" field of concepts at all? The determination of concepts by means of meaning postulates stays necessarily within one field; if that field and its internal structure are under debate, its postulates cannot be used to solve the conflict.[51]

To that purpose, relations between different fields of concepts have to be considered – relations that are closely associated with the concept »implementation« between abstract data structures. The concepts of one field are conceived of as a particular combination of the concepts from other fields – like data types that are understood as combinations of types from other abstract data structures. Let us consider two important distinctions: instances of *substantial* concepts carry properties or stand in relations that are expressed by *attributive* predications and need a substance to be carried by. A field of concepts is usually centered around a main substantial concept, and includes all the attributive concepts the instances of the main substantial concept have necessarily due to being such an instance. Triangles are necessarily planar, have three corners, and at least one of their inner angles must not be smaller than $\pi/6$. Triangles may also be necessarily either right or oblique, either equilateral, isosceles or scalene.[52] That a certain instance of »triangle« is, for example, made of iron does in fact happen; but it is not an attribute this object has due to its being a triangle. We have to distinguish the *essential* attributive concepts from those of *arbitrary* attributes or relations. In KL-ONE, essential attributes must be part of the T-BOX, arbitrary ones must not. The latter are not associated in a systematic way with the substantial concept in question: i.e., they are not part of the same field.

Arbitrary attributes occur if something that is currently viewed as an instance of the substantial concept of one field happens to be viewed additionally as an instance of the substantial concept of another field, e.g., a triangle as geometrical object and as a material object. In this case, the relation between the concepts of different fields is only mediated by a common instance (cf. Fig. 51).

[51] Nor can, of course, any reference to examples falling under the concepts in question help, since the opponent still rejects those concepts and firstly wants to be convinced of using them.

[52] Such ranges of mutually exclusive attribute values are called *incompatibility areas*.

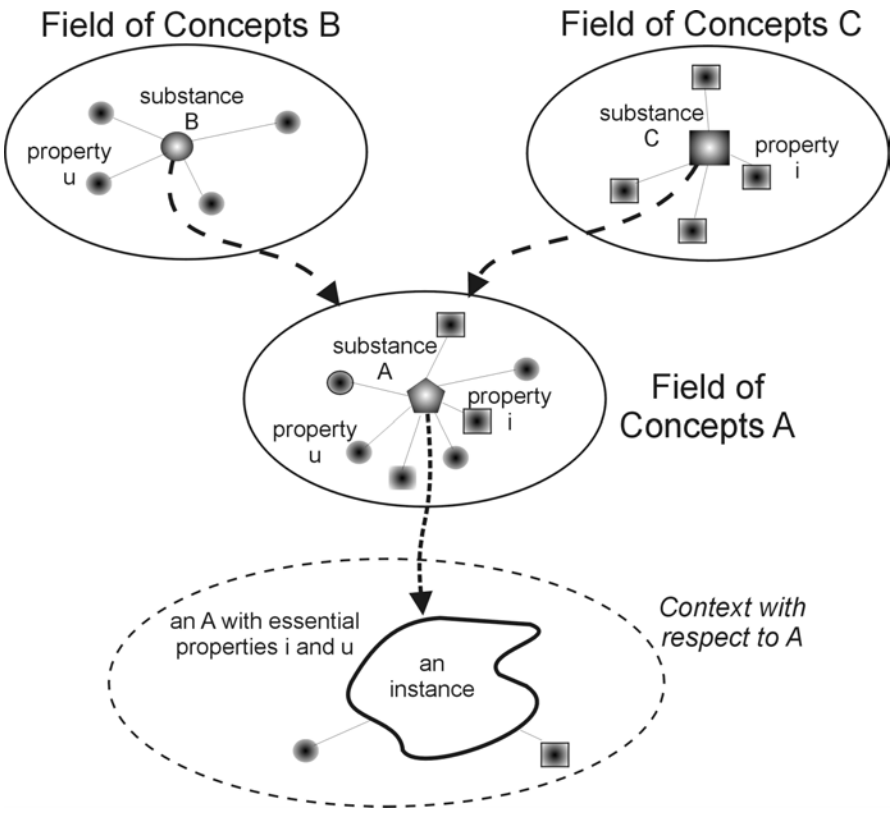


Figure 52: Graphical Schema on Field-External Relations ("Implementation") – the Argumentational Emergence of a New Kind of Entities

The "trick" of grounding the meaning postulates of a field of concepts in a rational argumentation now becomes obvious: we have to take those instances as instances of a new type of substantial concept that has necessarily the attributes formerly rated as arbitrary (cf. Figure 52). Note that the field-external relation then provides the new field with its instances: any concept of that field is a combination of the "habits of distinguishing" inherited from the other fields. The entities of that type have projections to entities of the combining fields; their attributes and relations are mixtures of the attributes of and relations between those projections. Nevertheless, it is important to understand that mentioning the field-external relations does not have an ontological meaning (about the things out there in the world). It rather introduces a particular argumentative strategy: the meaning postulates of a field of concepts are not seen as something that cannot be questioned any further ("that's how it is; you have to believe it"): something set by some untouchable authority. They can be understood as something constructed from other (usually simpler) sets of axioms following some construction schema, which may be discussed and changed by the community, as well. Only under that perspective, new kinds of objects "emerge".

For computer scientists, the analogy to abstract data structures may be easier to grasp. The algebraic specification of a data structure allows us to abstractly analyze complex data types and to define complicated algorithms, i.e., to argue whether or not certain structures are possible within the system. But the system itself does not tell us anything about how to find concrete instances (how to make the system "real") or about the rationality behind its axioms (does the system make "sense"?). An implementation, i.e., the systematic combination of several autonomous data structures, gives us indeed another type of rational arguments that allows us to show that the combined data types and

their implemented relations indeed follow exactly the given specification – or taken inversely: we can show that that specification is "realizable" and meaningful.

The field of concepts of sortal objects is quite complicated; it is almost impossible to understand its internal structure in its entirety, i.e., to explicitly know all of its meaning postulates. There are, in other words, only partial, incomplete specifications available for the rational argumentations about the structure of spatial objects – many aspects remain "intuitive" ([LEIBNIZ *1875, §24*]). However, it can be conceived of as "implemented" by (*i*) perceptible geometrical Gestalt concepts, and (*ii*) abstract entities that stand in meronomical relations with each other (cf. Sect. 3.4.4). This implementation schema gives us at hand a mechanism for rationally reconstructing the complete structure: we can refer to the geometric or meronomic projections in order to "found" the meaning postulates of the field of sortal objects.

In principle, those structures of rational argumentation as depicted in Figures 51 and 52 can also be translated to the distinctions in KL-ONE. Unfortunately, the family of KL-ONE knowledge representation formalisms does not (so far) include relations between different T-BOXES corresponding to the field-external implementation relation. But for our illustrative purpose it may be admissible to assume such a relation. That is, we assume a T-BOX governing the contexts with geometric Gestalt individuals, another one containing the rules that describe how to deal with parts and wholes, and a third one constructed (implemented) by the other two and determining how to deal with sortal objects and the contexts they appear in.

4.3.1.3 Reasoning with Spatial Objects

The effect of such a conceptual reconstruction of the field of spatial objects becomes quite clear when we look at *Spatial Resasoning*, the deduction schemata mediated by the meaning postulates of the spatial field in particular with respect to spatial relations. Spatial relations are attributive concepts of spatial objects, and their verbal appearance is mostly given by locative prepositions, like 'in', 'in front of', 'close to' or 'across'. In other words: the (intra-lexical) semantics of those prepositions is an explicit formulation of exactly the set of corresponding meaning postulates of the field of spatial objects.

This set includes statements about the transitivity or reflexivity of a relation or the (in-)compatibility between two relations. For example, we would usually agree in all naïvety that 'in' is a transitive relation: when a thing is in something else that is again in a third object, then the first thing is also in the third object. If something is to the left of another entity, it is also true that the latter entity is to the right of the first one, and *vice versa*. Similar to the example with Socrates' mortality given above, the deduction schemata used here combine several sentences about concrete instances (of spatial objects) with a meaning postulate that relates the predications (cf. Table 2).

A scrupulous empirical investigation leads to the insight that the meaning postulates of spatial relations are much more complicated. In some cases, »in« is transitive, and in others it is not, depending mostly on the types of spatial objects involved. For example, a pencil is in my hand, and my hand is in a glove – yet, 'the pencil is in the glove' cannot be deduced. The bee is in the rose, and the rose is in the vase – but nobody would expect to find the bee in the vase. In fact, no one could simply produce an exhaustive list of all the meaning postulates relevant for the "spatial T-BOX".[53] And even if some-

[53] Note that this ignorance does not hinder anybody's ability to deduce or to rate the correctness of spatial deductions. That remains however in most cases a purely intuitive skill.

Table 2: Two deduction schemas of the spatial (sortal) field of concepts

given empirical:	Your apple is in the bag.
given empirical	The bag is in my kitchen.
mediating conceptual:	"In" is transitive.
deduced empirical:	Your apple is in my kitchen.
given empirical:	The ball is right of the vase.
mediating conceptual	"Right" is the converse of "left".
deduced empirical:	The vase is left of the ball.

body could, the others just have to believe him; what could be their arguments beside their very private intuition.

The meaning postulates of spatial relations are certainly a central part of the data type »image content«: we have to use corresponding locative prepositions in order to describe the arrangement of spatial objects depicted. However, we do not really need a list of explicit meaning postulates. We can use a generative schema instead: the implementation schema of the spatial field. The approach of the French AI-group in Toulouse around ANDRÉE & MARIO BORILLO has demonstrated this method in great detail and with much success. In particular the study about 'dans' (the French version of 'in') by LAURE VIEU [*1991*] exemplifies how a very complex structure like the transitivity of »in« can be generated by two calculi that are combined systematically: a mereogeometry for the Gestalt aspects of objects and a meronomy for their part-whole aspects.

VIEU's overall schema is too complicated to be described here in any detail. It may suffice to mention that for certain kinds of spatial objects some part-whole relations are more relevant than others. Some types have also special geometric components. The material as the most typical and general part of a spatial object is usually considered as determining the geometric region relevant for being in that object: "The nail is in the wall". However for container objects, for example, their material forms only a secondary option: the largest part of the object's convex hull (beyond the geometric projection of the object's material) is the primary region for "being in that container". Essentially, »in« is only transitive if the part-whole-relations involved in the particular cases are compatible. Some pairs of »in«-instances are not transitive, because the types of part-whole relation involved there cannot be combined accordingly.

The main effect of this generating schema is indeed a shift of the level of explanation considered. The strange pattern of transitivity of "in" is not something invented rather arbitrarily by some ancient language creator. Nor does it simply follow the dark paths of individual intuition alone. We may view the deductive schemata of Table 2 like that, of course, if no dissent about the meaning postulates is to be solved at that time. However, if we need to motivate them we can change the perspective and generate the meaning postulates as emerging from the systematic interaction between the geometric projections of the parts of sortal objects. The deductive schemata appear then as a synthesis of deductive schemata of the constituent fields: a systematic mixture of elements from the geometric T-BOX and the meronomic T-BOX.[54]

A geometric model as used in computer graphics is essentially a more or less instantaneous three-dimensional geometric projection of a corresponding sortal object.

[54] We examine this point again and in more detail in Section 5.4.

Figure 53: Several Manners of Depicting Movement

Every geometric model can indeed be associated with many quite different types of picture contents: they may differ in their internal materials, or in the role the components have with respect to the identity criterion.

4.3.1.4 A Perspective on Perspectives

In consequence, »picture content« is associated to »picture« by a chain of two different projections: we have to consider (*i*) the projection from the complete sortal to an instantaneous geometric model in 3D, which (*ii*) must be projected into the 2D geometry of pictorial syntax. Only the combination of these two steps can be rightly adressed as the (inverse) content relation Rep^{-1}. Most astonishingly, theories on pictorial perspective often ignore the first step.

Naturalistic depiction style – in computer graphics usually called 'photo-realistic' – is reached if the following restrictions are applied: in step (*i*), the temporal dimension of the sortal objects involved is completely reduced to just one moment; and in step (*ii*), the resulting three-dimensional arrangement of Gestalts is geometrically projected to a two-dimensional subspace. The general principles of that transformation are known since the Renaissance era and correspond to a simplified version of physical optics ("ray optics"). Note that for this projection, the integration of all possible points of view, which is characteristic for the sortal field, is abandoned (cf. Sect. 3.4.4): one individual viewer perspective has to be specified. It originally defines the two-dimensional subspace of the image plane. Obviously, the apparent "simplicity" of the laws of the "natural" central perspective follows essentially from the fact that this secondary projection step is one completely within one field of concepts (geometry), while the primary step – often missing in discussions on pictorial semantics – refers to a much more complicated relation between different fields (cf. e.g., [REHKÄMPER *2002*]).

Naturalism is, of course, not the only option for the projection relation. Each of the two steps may be performed in alternative ways. First, the projection to the geometric constituents of the sortal objects needs not to focus on just one moment: the ability of sortal objects to move or change their shapes is characteristic for this concept. The pictorial representation of movement has a rather long tradition reaching from Australian aborigines, who bark paint the movements of their mythical ancestors by means of traces, i.e., sequences of their footprints (cf. again Fig. 36, p. 56), to MARCEL DUCHAMP's series "Nue descendant un escalier" of 1912/13 – giving just two examples.[55] In DUCHAMP's picture, deformations of the object's geometry also play an important role: in the course of the motion depicted, the parts of the human body change their shapes and their relative positions (of course without changing the identity of that sortal object). The art of depicting movement or other transformations has been perfected in the sketchy presentation styles of comics: motion blurring is often mixed with various

[55] As has been mentioned above, such pictures do at least partially leave the range of strict representational images. They include typical elements of structural pictures, and, in the case of the "Nue descandant un escalier" belong even to the category of reflective pictures with a representational core.

Figure 54: Isometric Representation

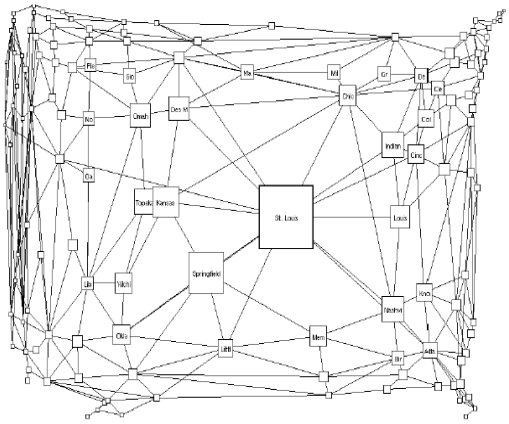


Figure 55: Example of the Fisheye Effect

forms of traces. Corresponding techniques can be adopted to computational visualistics: M. MASUCH [*2001, Sec. 7.2*], for example, has collected algorithms for calculating non-naturalistic images with multiple contour lines, motion lines (cf. Fig. 53) or motion blur from an animated geometric model.

There are alternatives of the second step, as well: multi-perspective images, for example. They are most prominent in art, in particular in Cubism (here again with the reflective momentum). However, even isometric pictorial presentations often employed in scientific visualizations and in some kinds of computer games (Fig. 54) may be counted to that group: they do not use central perspective but the closely related parallel perspective. No individual camera position is marked in this case (or the camera is thought of being positioned in infinity), so we can show with such a picture an integral view from many perspectives – or, if one prefers: from a God's eye view. The generalized point of view indicates that an overview is given that abstracts from individual perspectives.[56]

The inverse – a "hyper-individual" perspective – is also possible and useful: in order to embed a focused region of interest in a broader context, an irregular geometric projection is used which is often associated with the optical device called "fisheye lens". Thus, two perspectives are taken at once in a single picture: one from the distance for the contextual overview, and one (or several) from a point of view close to the interesting part of the spatial arrangement (Fig. 55). Typically for fisheye projections, the transition between the two areas corresponding to the different points of view is smooth (in contrast to the inset of a magnifying glass pixeme). In a generalized version proposed essentially by G.W. FURNAS [*1986*], several "focus points" – the parts of the representation that should be shown close – spread a "degree of interest" to all surrounding content elements that in turn determines the viewing distance with respect to that element. Indeed this generalized fisheye view with its multiple foci is particularly useful for abstract pictures and in interactive systems (cf. [PREIM *1998, Sec. 15*]). Maps, for example, and in particular city maps are often made with a fluctuating scale emphasizing important places while unimportant areas are diminished.

Many other forms of "non-photorealistic rendering" (NPR) in contemporary computer graphics have their conceptual bases in alternative projections from the complete sortal object to the geometric components of some of their parts, i.e., in the interaction

[56] It is also a little bit more easily to handle, which is an important criterion for the computer game scenario: when the gaze moves around a strategic map, for example, no variations in perspective distortion have to be calculated.

of the two steps. Concentration on contours, for example, may be viewed as a purely geometric operation. But, the sortal level determines which inner parts are important and must be present by some contour, as well.

4.3.2 Two Levels of Computer Vision: An Example

In the previous section, the projection relation has been in the focus of attention. Its internal division naturally plays a role, too, for its inverse relation Rep mapping the structural elements of a picture vehicle to the elements of »picture content«. In computational visualistics, this representation relation forms the central aspect of computer vision, and from there, also for the computational theory of visual perception in cognitive science. We come back to the connections between computer vision, visual perception, and picture understanding in the third part of this section, after having had a closer look on the components of the relation Rep.

Given a certain picture vehicle usually in the form of a pixel image, the task of computer vision is to find the corresponding pixemes and interpret them as a configuration of spatial objects. As a consequence of pictures being perceptoid signs of the visual sense, these steps are assumed of being closely related or even identical to corresponding cognitive mechanisms of human visual perception. Gestalt psychology has articulated the most convincing collection of grouping principles for the latter, and thus plays also a crucial role in computational visualistics for finding pixemes from a pixel matrix. METZGER [*1966*] lists seven main Gestalt factors:

1. Similarity: similar elements of the perceptual field ("tokens") tend to be grouped into a Gestalt
2. Proximity: tokens that are nearby tend to be grouped
3. Common Fate: tokens that have coherent motion tend to be grouped
4. Objective Attitude: new tokens tend to be grouped by means of the same principle that groups the older elements
5. Continuity: tokens that lead to uninterrupted, smooth curves tend to be grouped
6. Closure: tokens that may contribute to closed shapes tend to be grouped.
7. Completeness: all tokens are integrated in the Gestalt organization of the perceptual field

On a general level, two structurally different phases of subsequent processing can be distinguished: in the lower phase, the primary data is processed "bottom up" (data-driven): the results depend essentially on that data and grouping rules alone. In the higher phase, intermediate data is related to other sources of information, an integration, which is usually performed "top-down" (goal- or expectation-driven). It is easy to recognize in these two phases the construction of Geometric Gestalts (or pixemes), and the recognition of sortal objects, i.e., the reverse of the two projection steps discussed above. Let us leave for a moment the static images and consider image sequences; the influence of Gestalt principles can be demonstrated better in that case, which is also more natural with respect to human vision.

4.3.2.1 Constructing Visual Gestalts – Or Finding Pixemes

In the following, an example for the "perception" of spatial objects by motion is presented in some detail (cf. [SUNG *1988*], and [KOLLER *1992*]). In most computational approaches to visual object recognition, the signal of a video camera, i.e., essentially a

sequence of matrices of color values, stands at the beginning: a three-dimensional array of pixels. There is no other relation defined between any pixel but the relation to its immediate neighbors, which may or may not belong to the same pixeme. This neighborhood relation together with the marker value associated with each pixel are the only criteria to be used for the Gestalt grouping. Let us assume for simplicity that the marker values are only taken from the scalar intensity dimension.

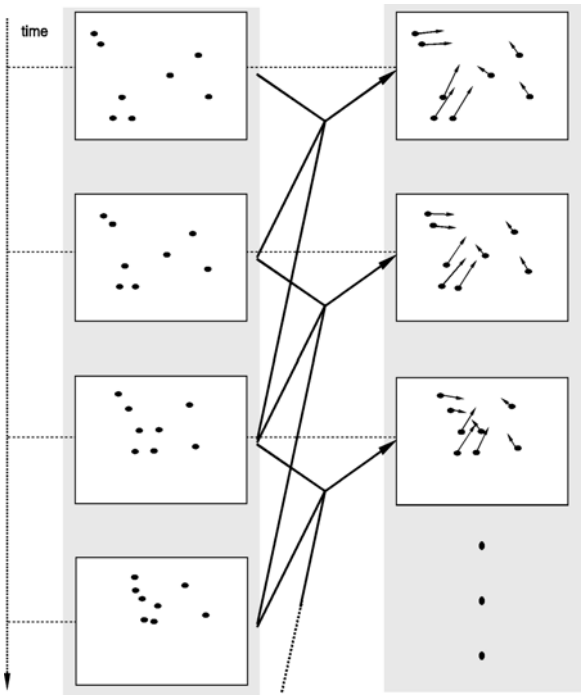


Figure 56: From Pixels to Vectors

The original amount of data is obviously quite high. Therefore in a first simplification step, some fields of pixels are concentrated to significant "features": the pixels with the local extremes of the intensity field. As anybody familiar with digital image processing knows well, filtering can easily do this. The concentration on the features is justified, as it is indeed a hidden coarse grouping step. The pixels in the vicinity are grouped together by the factors of similarity and objective attitude. However this grouping remains hidden until other Gestalt principles can be integrated: features are still only single instantaneous pixels that form the crystallization cores for further grouping factors.

To that purpose, the features at consecutive instants are grouped to instantaneous (velocity) vectors, if they are of the same kind (minimum or maximum) and at almost the same position (Fig. 56). Here quite obviously, the Gestalt factors of similarity, continuity, and objective attitude collaborate in forming new instances of a type quite distinct from the scalar intensity pixels. Still, the velocity vectors do not extend over time: they represent instantaneous local velocity.

In a third step, closely positioned similar vectors at one instant are grouped to spatially extended entities. If several of these still instantaneous entities happen to be close to each other and have a similar velocity vector, they are merged (Fig. 57). This originally leads to spatially extended entities: here finally the initially grouping hidden in the features is made explicit. Following the principle of wholeness, the field between the vectors grouped together is considered also as part of this new Gestalt: their convex hull defines the border.[57] The average velocity vector is calculated and taken as a property of these "object candidates" as they are called.[58]

A temporal extension is still missing; object candidates are pure two-dimensional geometric individuals. That is, they are the basic content-bearing pixemes in this example, inducing further pixemes by their borders and geometric arrangement by the internal rules of mereogeometry (cf. Sect. 4.2.2.1).

[57] In the schematic Figure 57, co-axial rounded rectangles have been used instead for simplicity.
[58] In computer vision, the expression "object" is unfortunately often used at this level already.

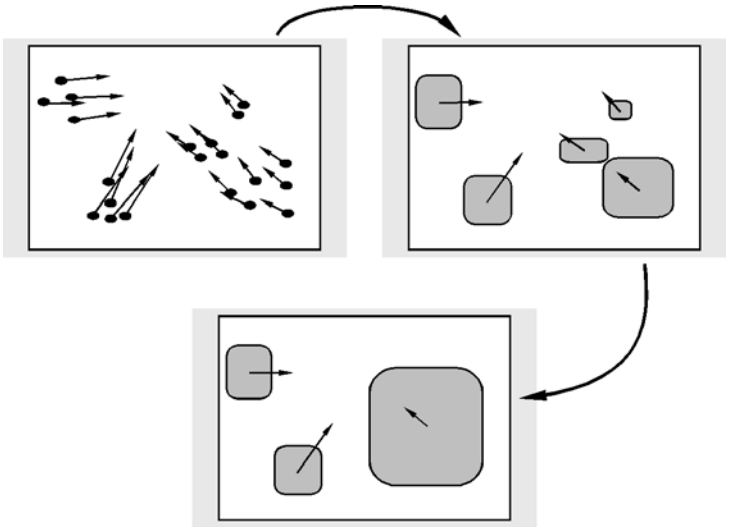


Figure 57: From Vectors to Object Candidates

Instead of using movement for the primary grouping step as in the example given, other systems of grouping factors can be employed based on static images. For example, abrupt local changes in intensity or color are integrated essentially by the factor of continuity into one-dimensional pixemes: edges. Furthermore, region-growing algorithms apply the factors of objective attitude, closure, and completeness, and expand initial feature pixels (accordingly chosen) to two-dimensional entities similar to object candidates based on color or texture parameters. The different methods can even be combined to form grouping synergies: motion-based object candidates may, for example, restrict the edge pixels taken into account for a contour grouping, or determine the starting points for region growing.

Let us come back for a moment to the motion example: an additional grouping can be used here. By means of the factor of common fate, the temporal development of object candidates is integrated into a new type of entity, called history fragments (Fig. 58). Note that history fragments do not include a criterion of identity restricting their shape: the temporally immediately neighboring object candidates grouped into one history fragment are of approximately the same size. But they may grow or shrink during the lifetime of the history fragment. Furthermore, history fragments may appear "out of

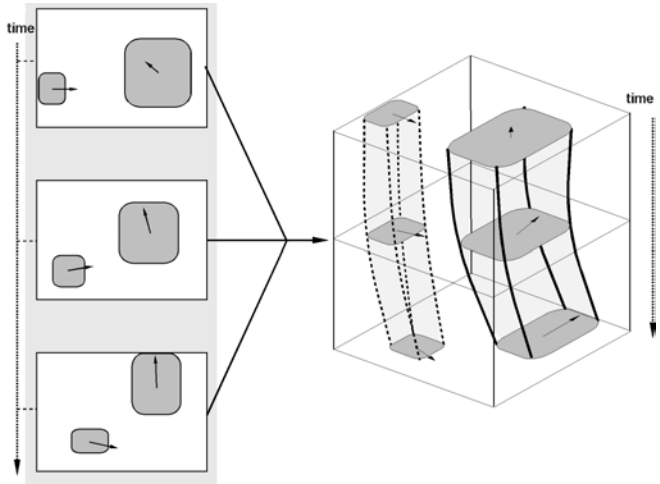


Figure 58: From Object Candidates to History Fragments

nothing", and they end when the corresponding sortal is occluded (even by blinking), or leaves the visual field. History fragments do not correspond to the histories of sortal objects. In fact, applying again Gestalt factors can repair some occlusions (cf. Fig. 59). However, the fusion of compatible history fragments already refers to a more complicated concept that has a criterion of identity with a more restrictive control of the possible development of shape, and of existence without immediate perception.

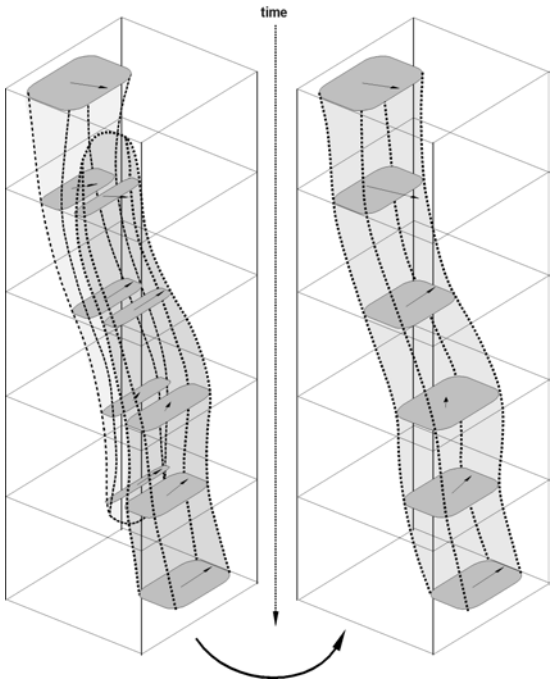


Figure 59: Repairing Occlusions: Fusion of Corresponding History Fragments

Object candidates are not yet spatial objects in the usual sense – as can be clearly seen in situations like that shown in Figure 60. Let us assume that two quite similar sortal objects approach each other, move closely together for a while, and then leave in different directions. The example sequence of grouping steps given above results in object candidates and history fragments as given in the schema on the left side of Figure 60. The history fragments are interrupted since the object candidates before and after the meeting are not similar enough for the temporal grouping: the common velocity during the meeting results in a "common" object candidate (C) much larger than the isolated ones (A, B, D, E). As has been mentioned before, the object candidates simply appear between frames "from thin air" or disappear without leaving a trace. Quite obviously it is impossible to reconstruct which of the original two candidates A or B has to be associated with which one at the end of the sequence D or E without additional knowledge.[59]

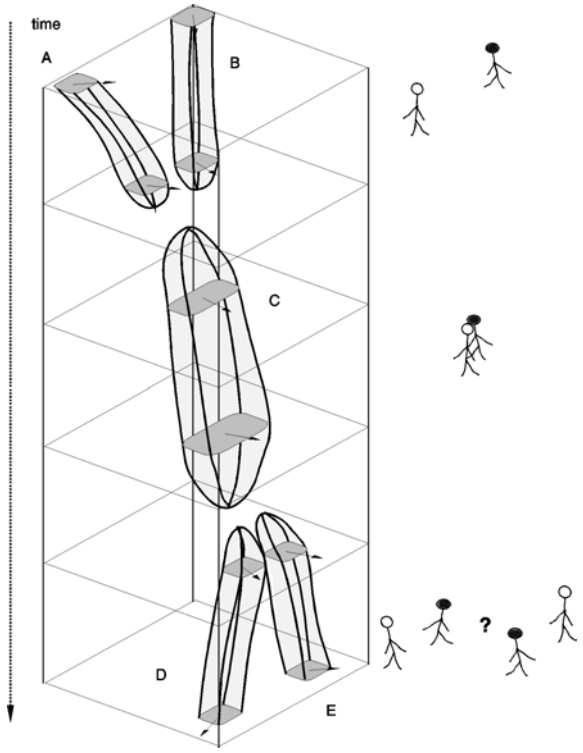


Figure 60: The Identity of Object Candidates

A similar problem appears in practice when human movements are "motion captured", be it in the context of movie special effects, com-

[59] The feature's marker values may help as indicated in Figure 60 by the color of the heads. If there are no such visual cues for identity beside the histories of the objects, the algorithm fails.

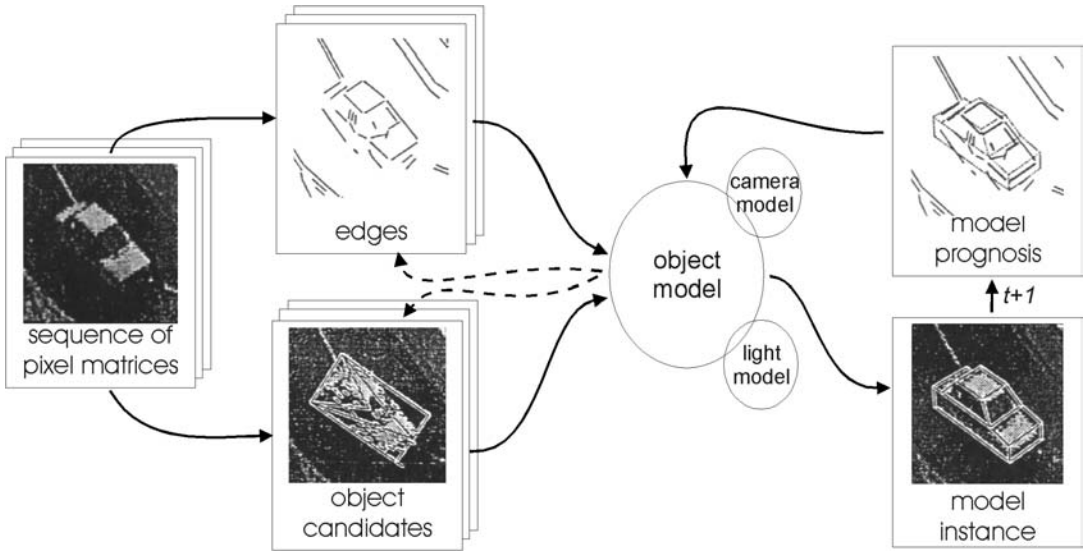


Figure 61: From Object Candidates to Instances of Sortal Objects

puter game design or sports teaching software systems: the pure "bottom up" grouping has to be complemented by additional "teleological" knowledge providing the conditions under which deformations, loss or exchange of parts and substance, occlusions, etc. do or do not alter the identity of an instance.

4.3.2.2 Instantiating Object Schemata

The established way in computer vision for adding such goal-driven knowledge is to employ "object models" (cf., e.g., [MARR *1982*]). They essentially describe which configurations of parts form an instance of a particular type of (sortal) object: they determine what deformations are within the range of that type of object, and (at least theoretically) what materials or parts play an identity-constituting role.[60] The projection of the object models to the object candidates found establishes finally the perception of spatio-temporally extended, persistent, and localizable entities – exactly the type of objects involved in spatial descriptions and realistic pictures. The largely geometric information about the *actual* configurations from the bottom-up phase is combined with information about part-whole relations governing the *possible* range of configurations.

Extending the example from the last section, Figure 61 sketches a possible procedure for that projection step. Let us assume that apart from object candidates by motion, edge fragments are computed in bottom-up manner, i.e., short pieces of straight lines where the intensity changes significantly in the pixel image. The top-down part controlled by the set of object models available works in a circular manner: a geometric projection of a model instance is adapted as close as possible to the edge segments within the corresponding object candidate thus establishing the instance at that moment. To that purpose, a camera model for the whole scene must be consistently instantiated. Note that this projection is a three-dimensional entity while edge fragments and object candidates still belong to the two-dimensional image space. The circle is completed by deducing an extrapolation of the present movement given by the velocity attribute of the object can-

[60] In most computer vision systems, much of this information is practically omitted since the reduced form suffices for their particular purposes – just as geometric models in computer graphics are highly incomplete but mostly sufficient substitutions for sortal concepts.

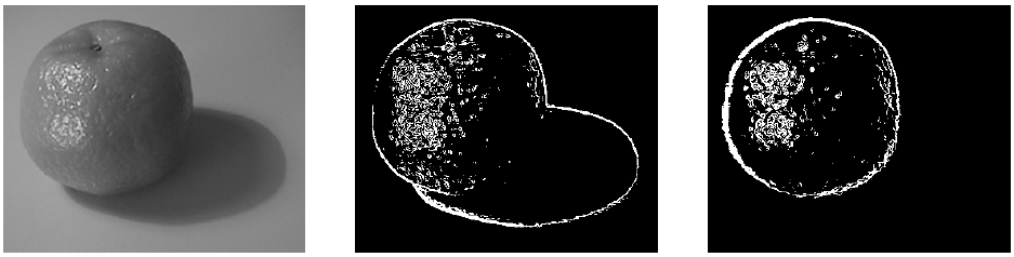


Figure 62: Distinguishing Shadows from Objects

didate. It is used as an expectation to help adapting the edge fragments for the next frame (hence the expression "expectation-driven").

Additionally, a light model can be added, which allows us to calculate shadows, as well (Fig. 62). Shadows are a specific consequence of the construction of sortal objects without being sortal objects themselves: for shadows (as for clouds) 1 + 1 may indeed not be 2 but 1 again. With the Gestalt-finding bottom up processes mentioned in the previous section, it is often impossible to distinguish shadows from objects. If the shadow touches the object, both are often included in the same object candidate even without many edge fragments between them. If the object is hovering so that its shadow is separated, two object candidates will result that appear completely unconnected. With the association of the object candidate with a sortal object in 3D space, the effect of light sources can be deduced and projected onto the image plane forming an expectation to be subtracted from the primary object candidates.

As a precondition for the (re-)construction of picture content, the deviations from central perspective should not be too pronounced. Isometric perspectives usually do not pose a problem since the perspective is locally not distinguishable from central perspective – one has to compare distant areas to note the difference. The cumulative effects of a locally small deviation from central perspective may add up to rather "screwed" perspectives as in M.C. ESCHER's "Prentententoonstelling" ("The Print Gallery", Fig. 63). It is the overall integration of the points of view in sortal objects that leads to the strange interpretation here: with every single glance, we recognize the Gestalts in focus as those of familiar sortal objects that are perhaps slightly distorted. But the integration of all the glances does not sum up to a consistent arrangement of those sortal objects in space.

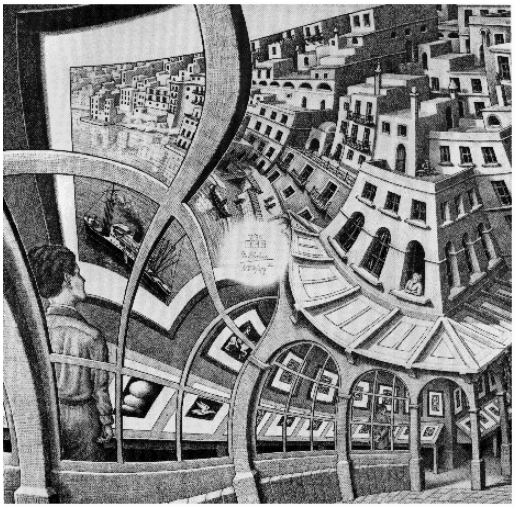


Figure 63: *Prentententoonstelling*

M.C. ESCHER *1956*

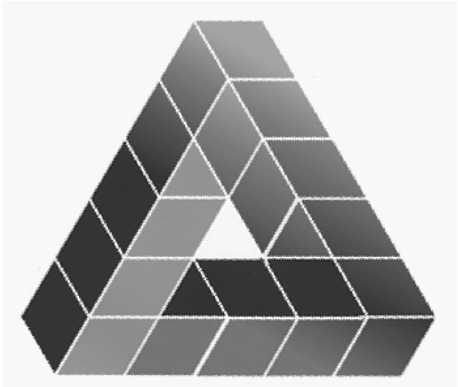


Figure 64: The "impossible" Penrose triangle

Figure 65: Sketch of One of the Physical Realizations of the PENROSE Triangle

This effect is most clearly demonstrated with a geometric figure called 'the Penrose triangle' (Fig. 64). Taken as a sortal object, the Penrose triangle seems geometrically completely impossible. In fact, there exist sortal objects that appear in pictures just like the seemingly impossible forms – if only viewed from one very special point of view (cf. Fig. 65). If the perspective is changed, the geometric illusion breaks down and a usual sortal object that is slightly strangely formed becomes visible. This indicates a further precondition for algorithms in computer vision: they usually fail if *special* points of view are taken so that (coarsely speaking) parts of sortal objects in different distances seem to merge – like the ends of the two arms of "PENROSE's object".

4.3.2.3 Determining Configurations

Since »picture content« must be more than a heap of sortal objects, a final interpretation step has to be mentioned: the perception of the spatial relations between the sortal objects. The concepts of such relations, which are usually articulated verbally by means of locative prepositions, form the basis of spatial reasoning, as has already been mentioned above (cf. Sect. 4.3.1.3). This reasoning is also important for the integration of the different local interpretations of the field of view into one unique understanding. Note that it is exactly the impossibility to ascribe correctly the relation between the viewer, the pictures and the gallery in ESCHER's "The Print Gallery" (Fig. 63) that disturbs the beholder here: What is "in" what? Which is "in front of" which? Etc.

The intensive linguistic investigation of locative prepositions in the last two decades (e.g., [HERSKOVITS *1986*] or [VIEU *1991*]) has made clear that the basing spatial concepts are not merely forms of geometric relations but a complicated system with geometric and part-whole aspects. They are proper attributive parts of the field of sortal objects. For "the teapot on the table", we have to concentrate on the geometric relation between (*i*) the underside of the bottom of the teapot's body, and (*ii*) a part of the surface of the table-board of the table.

Following HERSKOVITS' analysis, a mostly geometric core relation (the "ideal meaning") is adapted to apply to certain parts (which may indeed be a whole sequence of part-whole relations) of the sortal objects in question ("object idealizations") [HERSKOVITS *1986, 40*]:

> *In a particular use of a preposition, the ideal meaning may have been transferred to another relation, one that is in some way closely related* [sense shift]*; this new relation may in turn be only approximately true* [tolerance shift]. *Moreover, the objects related are mapped onto geometric objects (matching the categories specified for the arguments of the ideal meanings) by processes of geometric imagination, idealization and selection.*

The full complexity of the attribute system of spatial relations has not been transferred to computational linguistics so far. Computer scientists here still have to trade off

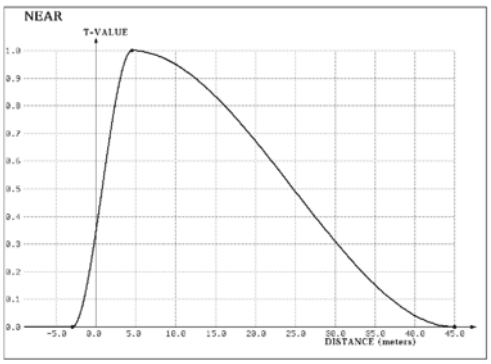


Figure 66: The "Ideal Meaning" of »Near«: Visualization of the Geometric Schema

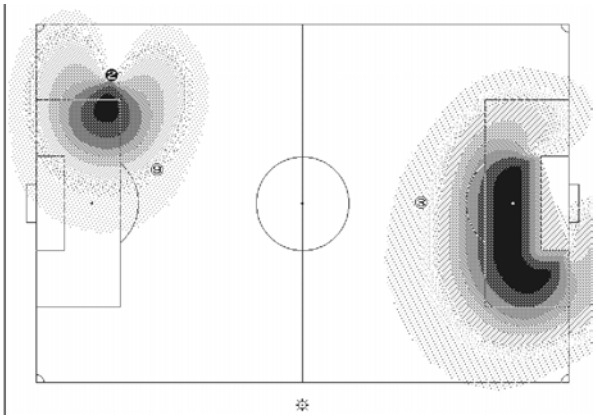


Figure 67: Two Adaptations of the »In front of«-Schema to Two Different (Geometric) Objects

between a detailed modeling of a very small subset of prepositions (cf., e.g., [VIEU *1991*]), or taking into account a larger set with a much more schematic treatment (cf., e.g., [HERZOG *ET AL. 1990*]). In the latter example, the ideal meanings are embodied by "applicability functions" that map "essential parameters" – abstract geometric attributes like »distance« or »direction« – to "applicability values", i.e., a form of fuzzy membership values in the interval [0.0, 1.0]. Figure 66 gives an impression for the dependency of the applicability value from the essential parameter »distance« for the core meaning of »near«. These schemata have to be differently adapted to the various types of objects by considering different parts as those relevant for determining the essential parameters. »Distance« for a point has to be calculated in another way than that for a line (taking geometrical examples just for the sake of simplicity) – the line is first idealized to the closest point. Figure 67 illustrates the corresponding results for the applicability function of »in front of« with the two essential parameters »distance« (analogous to »near«) and »relative direction« (approximately a Gausian curve with medians at 45° around 0° toward the object localized).[61] The "clouds" are a visualization of the applicability values for a zero-dimensional localized object for different positions (dark ≅ high value).

With the spatial relations, the spatial arrangements of objects in a picture can be completely classified. Thus, the primary »picture content« of representational pictures is finally determined. Further transpositions to higher fields of concepts, i.e., the one of intentionally acting creatures, may be based on this foundation. They are important for the practical use of most pictures and may even guide the construction of the sortal objects in a goal-driven manner. But those additional interpretation steps do not really contribute to the semantics of visual perceptoid signs *in the close sense*, i.e., to the relation between contextual pre-objective Gestalts and sortal objects. They are the same for the interpretation of a text describing the activities of human beings in a reduced manner by using spatial attributes alone.

4.3.2.4 Computer Vision and "Picture Understanding"

Most of the discussion so far seems to support an implicit identification of seeing, computer vision, and picture understanding, reflecting the approach in cognitive science

[61] The extrinsic form is shown, i.e. "in front of as seen from a point of view explicitly given". In Figure 66, those points of view are the left penalty spot (for being in front of a player idealized as a point), and the mark below the center line (for being in front of the right goal area).

(cf. [MARR *1982*]). One of the commonly used translations of 'computer vision' in German is indeed 'Bildverstehen' – i.e., literally: 'picture understanding' (as of human beings). The grouping of elementary pixemes into complex ones and their interpretation as sortal objects and their parts can serve as a model for human visual perception to some degree. We also may implement the algorithms with a computer connected to a video camera and receive a more or less stylized verbal description of "what the computer sees with its video camera" (cf. Sect. 5.4.3).

The sketches of argumentation-theoretic considerations above indicate an alternative and more precise approach. After all, the computer is still not an entity of the field of concepts needed to ascribe visual perception (cf. Sect. 3.3 & 3.4). The two steps of the representation relation described in the last section are indeed used to structure (parts of) the rational argumentations concerned with visual perception. The implementation has the purpose of exemplifying the argumentation in specific cases – and that is exactly how they are used when presented at a cognitive science conference. Note at this place that any computational realizations of semantic aspects of pictures are necessarily self-referential in the following sense: they are an implementation of an implementation relation since the field-external relation to the 3D-models already corresponds to an implementation of the concept of sortal objects as has been mentioned earlier.

At least the naïve identification of computer vision with picture understanding seems to be quite justified since the video camera apparently delivers pictures – but who for? (cf. again Sect. 3.5.3). »Seeing a scene« and »seeing a picture of a scene« are quite different concepts. Even »seeing the retina image of a scene« cannot be identified with »seeing a scene«. The second makes sense in everyday life in almost every context and does not involve pictures at all; the first happens frequently only in an ophthalmologist's office and does positively not concern the retina of the one doing that seeing. We rather keep this distinction when computationally modeling the concept »image«.

Indeed, the algorithms of computer vision deal firstly with the vehicles of pictures *for us*. This does, of course, not exclude that another picture vehicle is again part of the depicted scene: a separate sortal object with a more or less flat colored surface and a relatively clear border, a frame, separating the picture plane proper from the rest of that sortal. That is, there is perceptual space beyond the frame that does simply not belong to the picture plane. But beyond the space of pixels forming the input data for a computer vision system, there is no perceptual space at all. Correspondingly, a sortal object divided by the viewing frustum is notoriously hard to recognize for such a system.

Apart from those formal characteristics, the true interpretation as a picture needs additionally the ability to recognize the (potential) use of that vehicle as a sign.[62] A context much broader than usually available in computer vision is necessary to that purpose. The sign users and their intentions must be considered, as well. That is indeed, pragmatics has to take over here.

Nevertheless, for most practical purposes the naïve identification still leads to technically sufficient solutions: computer vision essentially deals with picture understanding not in all but in many relevant respects because visual perception plays such a dominant role for the data type »image«.

[62] Indeed, the transition corresponds to considering resemblance$_{\beta}$ instead of resemblance$_{\alpha}$.

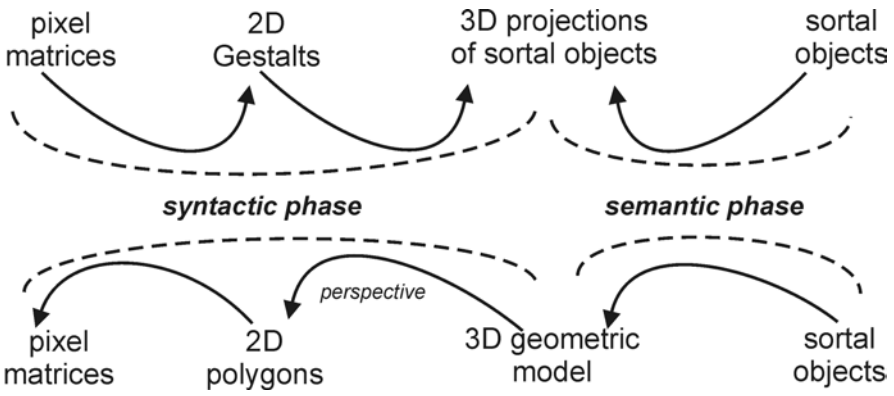


Figure 68: The Two Corresponding Phases of the Relations Rep and Rep^{-1}

4.3.2.5 Reference Semantics and Pictorial Reference

The two steps of the projection relation Rep^{-1} fit closely to the two steps of the content relation Rep (Fig. 68). The latter constists of a preparation phase that deals essentially with syntactic elements: the construction of visual Gestalts can be seen as "pixel parsing" in analogy to the parsing of letters to words and higher syntactic entities in linguistics. This segmentation phase corresponds roughly to the second phase of Rep^{-1}, the "rendering", where the pixels of the final image are calculated from the complete geometric Gestalts. The first phase of the projection relation is the projection from the sortal objects to the geometric 3D-models, which may be viewed as a preparation step for the rendering. It has its immediat pendant in the process of actually interpreting the pixemes as the geometric projection of sortal objects.

The previous sections could in fact be part of a linguistic treatise in the framework of reference semantics, as well. In the beginning of Section 4.3, reference semantics was characterized as the linguistic investigation interested in particular in the relation between words or sentences and the particular things or matters of affairs they are used to refer to. Reference semantics is especially interested in the study of verbal references to concrete instances; though, the extra-linguistic referents cannot be included *per se* in a rational argumentation. They are considered only by means of our perceptions of them and actions with them. An essential aspect of this relation is covered by the visual sense – it is equivalent to our relation Rep. In general, the various descriptions of what is seen constructed in the Gestalt-forming and -interpreting phases form distinct contexts (in the sense of Section 3.4.1) on different descriptive levels. Perception can in general be conceived of as the systematic relation between those contexts: the description of one level is used as the referential context for the description of the next higher level.

The contrast to intra-lexical semantics is therefore not really one between those linguists dealing only with words and those dealing with words and the world. The latter are essentially relating words, too. But while the former stay within one field of concepts (e.g., the one of sortal objects) using a semantic meta-language of that same field, the latter investigate the relations between several fields and are therefore able to use the additional schema of rational argumentation to formulate semantic relations. Figures 69 and 70 illustrate this usage. The reference semanticist can refer to the languages of the constituting fields *B* and *C* with respect to field *A* instead of the artificial meta-language, the internal structure of which is just as (un-)justified as that of field *A*. The fields *B* and

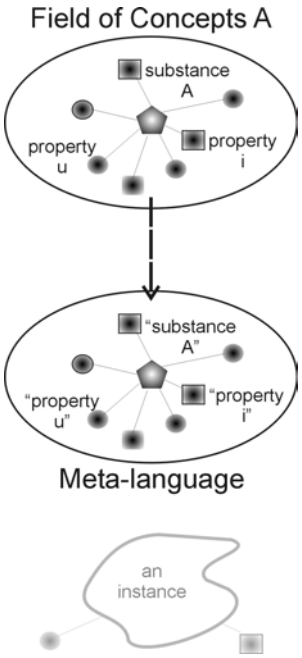


Figure 69: Intra-Lexical Semantics

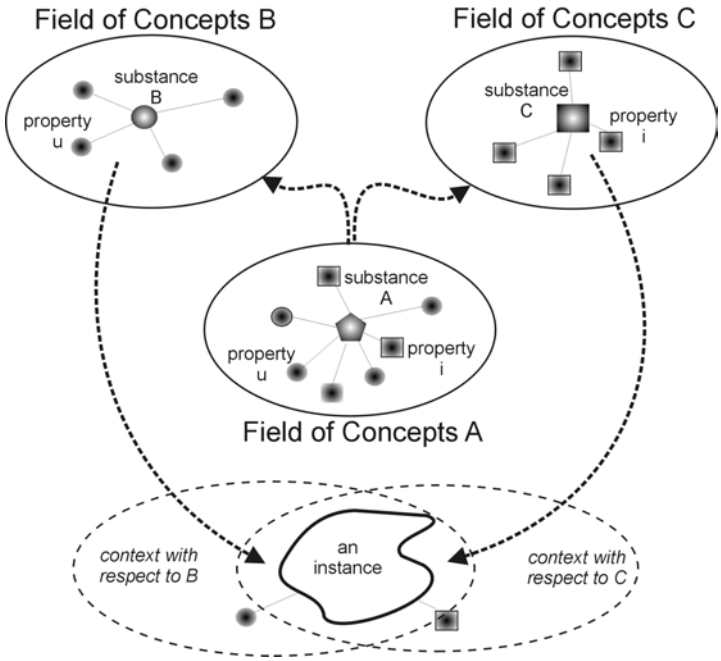


Figure 70: Reference Semantics and Field-External Relations

C do not simply copy the structure of *A*; they are instead employed to ground that very structure. But they also have genuine reference relations themselves, sensory-motor routines, which we can view as being "inherited" to the field constructed: they thus firstly constitute the reference relation for that kind of objects. That is: we see sortal objects (instances of field *A*) by means of perceiving Gestalt objects (or pre-objects, field *B*) and thinking simultaneously of coordinated part-whole relations determining the potential histories of the sortal objects (field *C*).[63]

That is, the problematic reference relation to extra-linguistic entities is essentially "shifted down one level" along the field-constituting relations in reference semantics. Note again: this is not an ontological shift but one in the structure of argumentations. It has consequences for pictorial reference, as well. So far, we have not dealt with the »picture referent« but only with »picture content«, i.e., the concepts behind the instances that originally allow us to see those instances at all, be that in reality or in the picture. Nevertheless, representational pictures do not (at least not primarily) refer to the concepts of sortal objects, although the latter constitute their content. The referents are still individual instances of sortal objects; instances, which in turn may serve to exemplify the concept in a secondary, metonymic level of reference.

Obviously, the "down-shift" that opens the way for the referents to enter the argumentation in linguistics cannot be used for the pictorial reference relation, as well. Take for example a picture of the Golden Gate Bridge: if one avows to be seeing something in that picture then either instances of Gestalt concepts are brought into the discussion or sortal concepts with an implication of their relation to Gestalt concepts. Only in the latter case is individuality possible beyond the actual context of perception because pure Gestalts do not have any identifying criterion with instances in other contexts if they are

[63] The same entity is viewed through a different pair of glasses, so to speak (cf. Sect. 4.3.1.2). We have to presuppose that those reference relations are not considered problematic, at least for that moment. They may, of course, come into the focus of a subsequent argumentation, as well.

not bound, like shadows, to sortal objects. As a referent, a geometric individual is over-specific; it is exactly the syntactic pixeme in that picture, and nothing more.

On the other hand, two cases can be distinguished for a sortal referent: the referent may be an unspecific individual object just introduced for the sake of some embedding communicative act; or it can be an instance we already know from other contexts – just a large red suspension bridge over some water spanning between hills; or the one and unique Golden Gate Bridge. In any case, the pictorial reference relation to the sortal instance cannot use the direct reference inherent to the geometric field: this leads just to visual pre-objects while the part-whole entities coordinated in the linguistic case are missing completely. It must take the inverse path through the sortal interpretation, i.e., the picture content, which then also provides the relevant meronomic components (Fig. 71). Note that this inverse use of the constitution relation of the field of sortal objects is the reason for the strange fact that the direction of the semantic projection in computer vision is reversed – expectation-driven – with respect to the overall "bottom-up" direction of that algorithm (Fig. 67, upper part).

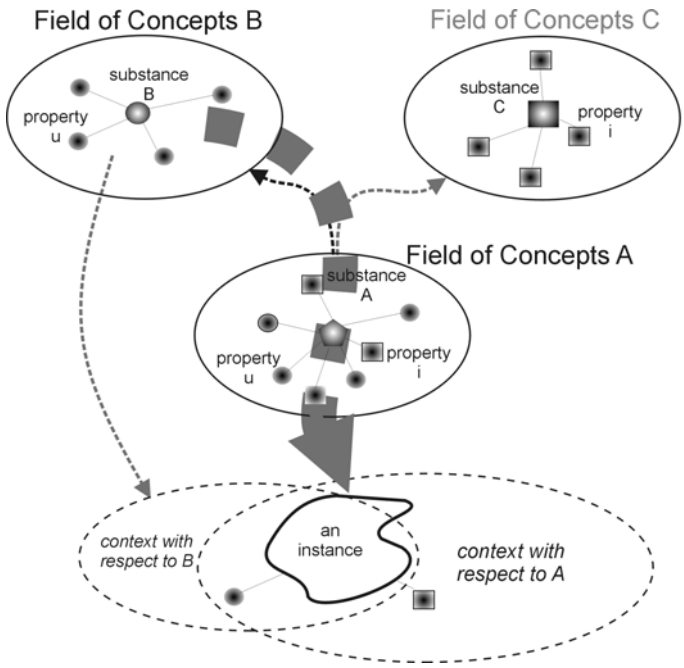


Figure 71: Reference Semantics for Pictorial Reference: Using the Path through the Sortal Field

There are three cases of pictures that are presented simultaneously with other contexts containing the same individuals. They are of particular interest for pictorial reference:

(a) the picture could be one of a set or sequence with other pictures meant to show the same sortal object;

(b) the picture is presented together with text that is used to refer to the same sortal individual (e.g., a caption);

(c) the picture is presented in the presence of the sortal individual depicted (as is often the case for pictures in instruction manuals; cf. also the passport example mentioned above).

The first two items are typical cases of co-referential sign acts. In the propositional sign system, there are sets of special signs for an easy use of co-reference – in particular the pronouns ("it") and anaphoric indicators ("that"). The pictorial sign system does not offer a similar tool. A set of the visual characteristics of the corresponding sortal concept has to be repeated in all the pictures of the sequence comparable to a (deictic) description. Each picture opens a distinct context that provides implicitly the deictic component. The identification of those individuals depends mostly on conventions and is not directly implied by the pictures.

In the third case mentioned, co-reference does not play a role. Nevertheless the problem of pictorial reference, i.e., to be able to identify the occurrences of one individual in two contexts, remains for the beholder. In the case of a repair manual for a laptop, for example, the pixeme activating the sortal concept of that laptop might also refer to a second individual laptop of that type necessary for performing the repair. Quite obviously, the interpretation of the pixemes, i.e., the picture content *per se* cannot indicate just that particular individual present in the situational context or referred to in the other picture or the text. Although controlled by the picture content, the pictorial reference as such remains unspecific since the picture simultaneously introduces the context for that reference. Verbal nominations, in contrast, refer relative to a context that is separately given. When referring to an individual purely pictorially we do that always without connecting that individual explicitly to other contexts. When referring by means of a nomination, we do so always by explicitly connecting at least two contexts.

Correspondingly, it is usually the text that is employed to establish the co-reference with a pixeme: if the descriptive part of a nomination in the caption fits to one of the sortal concepts in the picture content we take it that the text refers to the same individual as that pixeme (and hence implicitly *vice versa*).

An approach for pictorial reference employing a purely symbolic identification (parallel to the entities of the intra-lexical meta-language) may serve as a simple solution, but non-semantic factors form an essential component if the data type »picture referent« is to be dealt with: e.g., the likelihood of the communication partners to know that individual object in other contexts and consider it being relevant for the communication situation at hand. At least in this respect, the concept »image« is indeed not too different from verbal signs: both depend on pragmatic principles for identifying individuals.

4.3.3 Embedding Semantics in Pragmatics

In summary, the full data structure basing computational visualistics must contain beside its syntactic component »picture vehicle« two more associated data types explicitly covering semantic aspects: »picture content« and »picture referent«. The relations between »image« and »picture content« have a syntactic and a semantic part; the first is effectively associated with structures of the image's »picture vehicle«, the second depends on a field-external relation relevant for the concepts of sortal objects that form the central aspect of »picture content«. This object constitution also links the data type »image« simultaneously to visual perception and to verbal language. The later connection is essential for relating »picture content« with »picture referent«, though pragmatic factors definitely dominate that type of semantic relation.

The investigation of »picture content« including the short sketch on »picture referent« has been oriented at the verbal transcription of the corresponding entities. While the contents are associated to predicative verbal phrases, the referents are circumscribed by definite descriptions, i.e., nominatorically used noun phrases. This seems to contradict the central finding of image theory sketched in Chapter 3 assigning pictures the main function of context builders.

In accord with the twofold medial use of contexts as the anchoring ground for a proposition, and as the result of interpreting a proposition, context building is linked to the pragmatic embedding of propositional sign acts more closely than the other partial sign acts. This double nature of contexts is immediately reflected by the two grammati-

cal Gestalts possible for perceptive verbs, e.g., 'seeing':[64] "I am seeing two shepherds, a lawn and a tree" (nominatorical), and "I am seeing that a man embraces a woman" (propositional). Viewed as sets of entities, contexts may be sufficiently approximated by the set of nominal phrases identifying those objects that are provided by the context. Viewed, on the other hand, as a structural whole in which the objects only form crystallization cores connecting this context with others, a context may also be adequately approximated by a more or less complicated proposition (in this sense, a novel is a context builder, too). As perceptoid context builders, pictures qualify for the two approximations, as well: depending on the task at hand, their meaning may be given by means of the referents accessible in the context they build; or as a complete set of propositions describing the state of affairs. The referents relate the context to other contexts with the same objects; the state of affairs distinguishes the context from others containing the same objects.

It seems from this point of view that the predicative component dominates both aspects. The nominatoric identification rests on the habits of distinguishing. Those habits also determine the transformation of a context by integrating a proposition (i.e., an additional predication). Correspondingly, the predicative nature of »picture content« seems to be more important on the preceding pages than the abstract function of context building. Though *within* one context, those habits of distinguishing need not already be concepts, i.e., corresponding to predication. Detectors and the associated sensory-motor test routines (corresponding verbally to quasi-predicates) are sufficient (cf. Sect. 3.4). Concepts and the full use of predication already depend on the ability of context-building. The preference for the predicative characterization of »picture content« is, thus, a simplification necessary to grasp semantic aspects of pictures at all.

Nevertheless, the true nature of contexts and the full meaning of context builders like images enclose always both aspects: the anchoring for a successive predication, and the result of previous predications. They are differentiated in the course of the surrounding activities, and the complete communicative setting must be considered if an adequate treatment is intended. Quite obviously, this setting includes more than one participant, and the role the pictorial sign act has for their other activities is the key for completing the generic data structure around »image«.

4.4 Pragmatic Aspects

The field of pragmatics has been characterized in the beginning of this chapter as the investigation of the complex formed by a communication act and the other related behavior, i.e., the embedding of the sign act in the "living practice" of the sign users. Indeed, semantics in the traditional sense – i.e., an investigation apart from pragmatics and restricted to those relations between sign vehicle and sign meaning that are independent from the sign act, its participants, and their further behavior – must remain relatively sterile. Even the transcription of meaning components into verbal expressions used above relates the picture use to another sign *behavior*. Furthermore, a valid theory of resemblance can be reasonably founded only with respect to the behaviors of those experiencing a similarity, as has been sketched in Chapter 3. If resemblance is taken as a basic ingredient of pictorial semantics, semantic considerations are necessarily "contaminated" with the pragmatic perspective.

[64] Obviously, a situational context is here introduced by means of the perception verb.

Correspondingly, a broader conception of semantics has already been used throughout the preceding section: that conception views semantics as a part of pragmatics, and more precisely, as the part focusing on what is mentioned as being signified by means of the picture vehicle by the sign user. The relations to spatial reasoning and argumentation theory mentioned in that context clearly demonstrate this shift of perspective.

We might think of classical semantics as the reduced pragmatics of a soliloquy (rather than the "God's eye view" semanticists have often supported). Since it is not the picture (or more generally: the sign) that shows or represents something: it is the picture (or sign) user who shows or represents something with the sign. In soliloquy, one directs and keeps one's own focus of attention by means of the sign on something that is usually not actually present. The discussion in Chapter 3 indicates that signs of the level of communication that includes propositions are in fact the only tools we have for performing such a peculiar behavior.

However, those signs are more generally used to communicate with somebody else: to direct and keep the focus of attention of a communicative partner on something. That is in particular: for allowing her or him to perform some behavior linked with the signified entity or to coordinate such a behavior with behaviors of the sender of the message. The net of relations between the sign act and those behaviors is meant by "the embedding of the sign act in the living practice of the sign users"; it forms the focus of interest of pragmatics.

Thus, the situational settings of the sign uses play a prominent role for pragmatic investigations. For computational visualistics, most of the traditional settings for picture uses are relevant, too. But the truly specific pragmatic setting is the use of pictures in interactive systems, to which we now turn first (4.4.1). An important tool for dealing with pragmatic aspects of interactive pictures is the anticipation of the potential beholders (4.4.2). But the sender has to be considered explicitly as well when the communicative authenticity of a sign act with interactive pictures is to be assured over and above the weak form of technical authenticity that can be provided by the medium as such (4.4.3). The rhetoric of structural pictures (4.4.4) and pragmatic aspects of computer art with its link to reflective pictures (4.4.5) complement our discussion of pragmatics in computational visualistics.

4.4.1 Interactive Systems as a New Type of Media

A well-known classification system of media theory [PROSS *1972*] distinguishes three types of media: whereas primary media (or media of class *I*) do not involve any technical devices that open the possibility of temporally or spatially separating the communicative partners, secondary media (or media of class *II*), like books or letters, involve devices on the producers side. If the communication depends on the use of special devices on both sides of the communication channel, a tertiary medium (or medium of class *III*), like TV or telephone, is used. Quite obviously, we can easily decide as a symptom whether technical devices are applied for receiving and/or sending. But are those symptoms already the true criteria underlying the appropriate classification of media intended? Is it not remarkable that primary media have as their precondition that all participating communicative partners must share the same situational context, while the sender generates persistent sign vehicles with secondary media that can be used for communication across temporal separation? And that media of class *III* enable the interlocutors to communicate across large spatial distances without significant loss of time? Those differences in the situational setting of the

communicative act must have an important influence on the content and form of a corresponding message.

There is a small amount of primary media uses of pictures, like a quickly drawn structural sketch employed in a verbal argumentation and thrown away afterwards; or the ceremonial sand paintings of Australian aborigines, the vehicles of which are also destroyed immediately after the ceremony. We may also count the showing of the picture in a passport for personal identification to this category. However, pictures are traditionally used mostly as media of class *II*. Their production is often a more or less complicated procedure that makes it impossible for most cases to generate a picture spontaneously like a verbal utterance, or to anticipate the need of a particular sign vehicle in advance, as in the case of the passport. More importantly, pictures are usually intended to persist over a considerable amount of time in order to allow the sign user to establish, as with written text, a communicative link between different times. This link may connect sender and receiver in the same person, or in the form of different sign users, forming a kind of external memory or a true act of interpersonal communication, respectively. Pictures in an exhibition or in a book are typical examples of images used in secondary media; so are films. Note that it is essentially the temporal separation of senders and receivers that determines those examples. Spatial distances between the situational contexts of the communicative partners that may also appear with media of this class are mere by-products. Overcoming the gap is a process that consumes a lot of time compared to the actual communicative activities.

In contrast to that, the transportation time of the message is almost negligible compared with the duration of the communicative act for media of class *III*. For spatial contexts far apart of each other, this can only be reached if *all* the interlocutors employ technical devices. Note that the fast transportation of the sign vehicles between the spatially separated situational contexts of the interlocutors is the major precondition for a two-way communication similar to the direct social interaction in primary media. Of course, pictures are also used extensively in media of class *III*: from sending facsimiles by telephone lines to the solitary tele-sensoring by means of surveillance cameras, as well as from digital photographs taken and sent by trendy mobile phones to the mass participation of viewing a soccer game by means of a life broadcasting in TV.

The construction process for representational computer graphics indicates clearly that a medium of at least class *II* is considered. The general structure of the device to be used for production has already been mentioned in the preceding section: a three-dimensional geometric model is provided by the computational visualist as the input data for a program that calculates a projection of the geometric model onto a two-dimensional image plane. The geometric model is a formalized description based on a data structure that allows the computational visualist to describe three-dimensional geometric Gestalts: the description of an individual's geometric and optical properties concentrates on certain aspects of the actual sortal object described (be it real or fictional). The projection creates another description based on two-dimensional matrices of elementary regions with color attributes (pixels). It is a certain presentation of a pixel matrix by a monitor, a projecting device or a printer that can finally be employed as an image.

The picture vehicle generated by the computational visualist could be used in the same way as a picture vehicle produced in the traditional manner, i.e., independent from the production process, as a perceptoid sign in a true communicative situation or in a hidden auto-communicative situation, a kind of pictorial soliloquy, that is. The printout

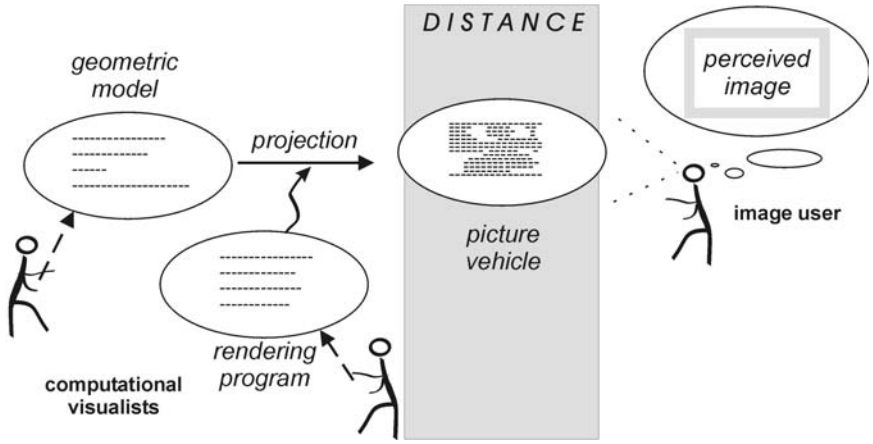


Figure 72: Presentation of Computer-Generated Pictures: The Direct Use

or the projection can be employed in many sign acts that are not at all related to the situation of the production.

However, the final step, i.e., the printing or projecting of the pixel matrix, is usually not considered of as being part of the production proper. The site of production of the pixel matrix and the site of its projection into a directly perceptible form can be (and often are) far apart from each other. Since the final presentation must be performed by means of another technical device on the recipient's side computer graphics have indeed to be conceived of as a typical medium of class *III*. In this case, the computational visualist who has provided model and rendering parameters is usually viewed as the primary sender of the sign act of the computer-generated picture (Fig. 72).

4.4.1.1 Media of Class *IV*

When dealing with computer graphics in interactive systems, the schema given in Figure 72 has, however, to be adapted in a particular manner: although the picture is still produced by means of the rendering algorithm from a geometric model, this happens at some point in time and place apart from the person to be considered as the primary sender in this communication, the one providing the model and the rendering algorithm (Fig. 73).

Take for example a textbook on human anatomy and its interactive pendant. In the book and in the interactive version, pictures illustrating anatomic objects, some of their relations, and some of their attributes are offered. The standard situation of use appears as a (pictorial) soliloquy: for example, a student uses the pictorial sign for focusing his attention on those anatomic matters in order to learn them. Or a physician wants to refresh her memory by means of showing that sign to herself. Although acting as sender and receiver simultaneously, the student and the physician have to trust the original picture producer and the technical devices transporting the sign vehicle to them. Otherwise the picture cannot be employed in an authentic soliloquial sign act.

For the traditionally printed textbook, this trust is essentially established by means of the social institution of the initial production process: the produced picture is persistent; it usually does not change significantly. This attribute is also viewed as a disadvantage of the traditional medium, which is finally cured by the interactive version. The users of an interactive textbook are not restricted to static, pre-fabricated images anymore; they can easily chose other perspectives, turn, scale, move or remove parts of the anatomic

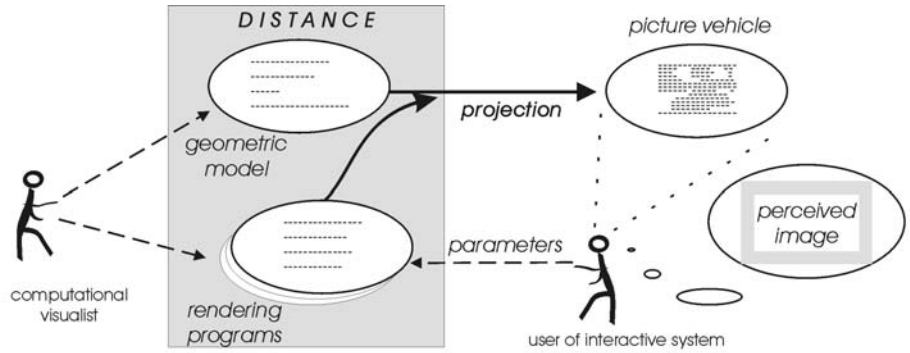


Figure 73: Computer-Generated Pictures in Interactive Systems: Tele-Rendering

objects displayed, zoom in or out, and even change the style of the presentation. To that purpose, the image is rendered at presentation time.

In consequence, the situational setting of sign production and sign reception seemingly merge – almost like for primary media; the appearance of a pictorial soliloquy is even stronger for interactive media, on first view. The essential pragmatic question is then: how can those sign acts gain authenticity? Of course, it is still the computational visualist who provides the model and the rendering algorithm that have to be transported to the users in order to generate the picture on demand. So the soliloquial use of the picture in the interactive textbook on anatomy is still a derived sign act borrowing its authenticity from an underlying sign act from computational visualist to system user.

We may use the expression "tele-rendering" for the situational separation between the preparatory design activities of the computational visualist and the actual image production that is finally induced by the users of the interactive system. Note that computer graphics does not necessarily imply tele-rendering although it has opened the way for the latter: computer graphics' potential to easily change the model or the style of rendering provides a significant variability of rhetoric elements adaptable to individual communicative contexts.

There is a profound difference to the other examples of media connecting separated contexts: whereas the unit to be transferred by technical means through space and time with secondary or tertiary media is formed by one single message, or more precisely one unique sign vehicle, tele-rendering can legitimately be viewed as transferring whole classes of messages/sign vehicles. Depending on the user's interaction, one of the instances of that set is realized in a particular user session. Tele-rendering therefore belongs to a different class of media altogether. We suggest calling this type "media of class *IV*" (or quaternary media) . The automatic production of verbal signs by language generation systems resulting AI research forms another member of that class. It is no accident that such programs are a main component of interactive systems, as well.

We have to expect particular consequences for the communicational function of any signs used in class *IV* media, and especially for the pictures created by tele-rendering. The rhetoric force of each concrete picture generated for a specific user must be carefully adapted by the interactive system to the particular communicational setting at hand if miscommunication with potentially fatal consequences is to be prevented: imagine, for example, again the interactive textbook in medicine, and the effects an insufficient act of pictorial communication can have in this domain.

Most of the investigations in tele-rendering so far investigate pictures in interactive systems in analogy to propositional utterances and their logical parts. A short overview

about their solutions is given in the following sections, leading to another component for our generic data structure: the beholder models. In those studies, pictures are usually employed in quite specific ways: as nominations or predications, the complement of which is mostly given verbally. The common background of the picture's most general function as a context builder is not taken into prominent view. In consequence, a proposal to adapt beholder modeling to context building is given in section 4.4.2.

4.4.1.2 The Selection Problems: Content

Language generation systems have a tradition longer than that of tele-rendering; some aspects relevant for the latter can be derived from corresponding AI research. Selecting the content ("what to say") and determining the form of an verbal utterance ("how to say") are distinguished on the general level and usually form separate components in language generation systems. For producing a concrete example, the two components usually have to interact.

An analogous distinction can be used for the autonomous generation of a representational picture in an interactive system – we may approximately speak of "what to show" and "how to show". Determining the "what to show" is on first view quite similar to selecting the content for a verbal proposition, i.e., which state of affairs is told about what objects. More precisely, this seems to be a completely semantic task. But of course, (*i*) not all verbal utterances are propositions (although many of them have propositional cores), and (*ii*) pictures are not really analogous to propositions. Let us deal with the second restriction first: choosing "what to show" must essentially be determining which context is to be built by means of the pictorial sign act. This can only be planned indirectly and depends on the perspective on the context: what objects are to be identified for the interlocutor by means of which attributes? Or: which stories are to be evoked, i.e., which states of affairs are to be shared? That is, we here meet again the propositional and nominatorical aspects of contexts we have covered exactly by the semantic aspects dealt with in the preceding section. This semantic core of a planned pictorial act may be embedded as partial sign act in higher level communicative acts, like propositions being used as parts of promises, requests, commands, and other speech acts – we shall come back to that aspect (mentioned first above) in the next section.

Let us have at this place a quick look back at one of the formalisms developed in AI for dealing with the content and referents of verbal utterances: KL-ONE. The intralexical conceptual rules of a field of concepts – its meaning postulates – are covered by propositions in T-BOXes while empirical propositions are collected into A-BOXes. That is, an A-BOX is the KL-ONE equivalent of a context. Essentially, the differences between an A-BOX representing the recipient's point of view and an A-BOX of the sender's focus of interest determine what objects can act as anchor points for nomination, and which attributes or relations are not shared yet and have to be communicated as predication. The process stays essentially within one field of concepts: definitions of complex concepts may be intra-lexically analyzed. Field-external relations are usually not considered.

For pictures, this schema is only partially applicable. Determining the context to be built by a representational image is certainly the main step, as long as the field of sortal objects governs this context. But this is not yet the »image content« we need. In the preceding sections, we have characterized »picture content« as the concepts involved in recognizing something in the picture space: that is, from the perspective of reference semantics, viewing the concepts together with their visual test routines inherited from

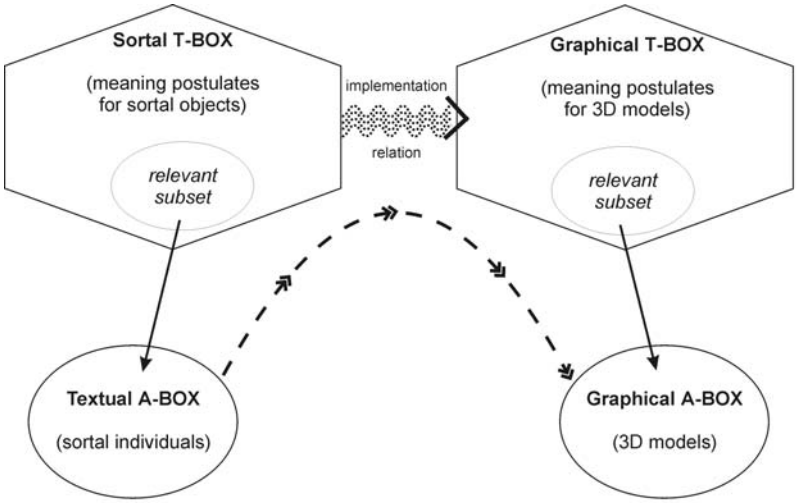


Figure 74: Selecting What to Show – from A-BOX to T-BOX and beyond

the implementing Gestalt concepts. In order to determine the »picture content« proper from an A-BOX describing the context to be build, we first have to extract the relevant concepts. We find them in the corresponding T-BOX where they are only given in their intra-lexical form. KL-ONE-like systems do usually not deal with field-external relations between T-BOXes or the inheritance of sensory-motor test routines determining the reference relations. However, we may assume that a corresponding relation between the concepts of sortal objects and their geometric projections – geometric 3D individuals – is given. In any case, the complex formed by the sortal objects derived from the context (i.e., an extract of the T-BOX governing the A-BOX selected) in their relation to corresponding geometric objects (i.e., relation between two T-BOXes) is exactly the instance of »picture content« we need as the result of selecting "what to show" (cf. Fig. 74).

Of course, in the typical situation of an interactive system, like the digital textbook on anatomy mentioned above, many of the pictures presented are not generated out of nothing (so to speak): they are essentially transformations of the picture shown the moment before. Or more precisely, it is a content already selected that is merely transformed leading to corresponding syntactic changes. On the level of the knowledge representation system, this corresponds to a given A-BOX to which certain propositions are added while others are deleted (since they are now irrelevant). Note that adding propositions is the only way of shifting the focus to new objects not included in the older context. The new A-BOX also has a different T-BOX projection determining the change in content.

An impressive example is given by the system *TextIllustrator*, an experimental interactive textbook on anatomy [SCHLECHTWEG & WAGNER *1998*]. Although not directly using a knowledge base, it allows a user (among other things) to change the image displayed on the left side of the screen indirectly by scrolling the text shown on the right side (cf. Fig. 75). The image always corresponds to the part of the text visible. Furthermore, clicking an expression marked in the text – essentially, those are the Latin medical terms – results in highlighting the corresponding object and eventually even in turning the scene so that the object can be clearly seen. While this latter effects are mainly part of the "how to show", and we shall come back to them in a minute, the "what to show" aspect is more dominant in the first function.

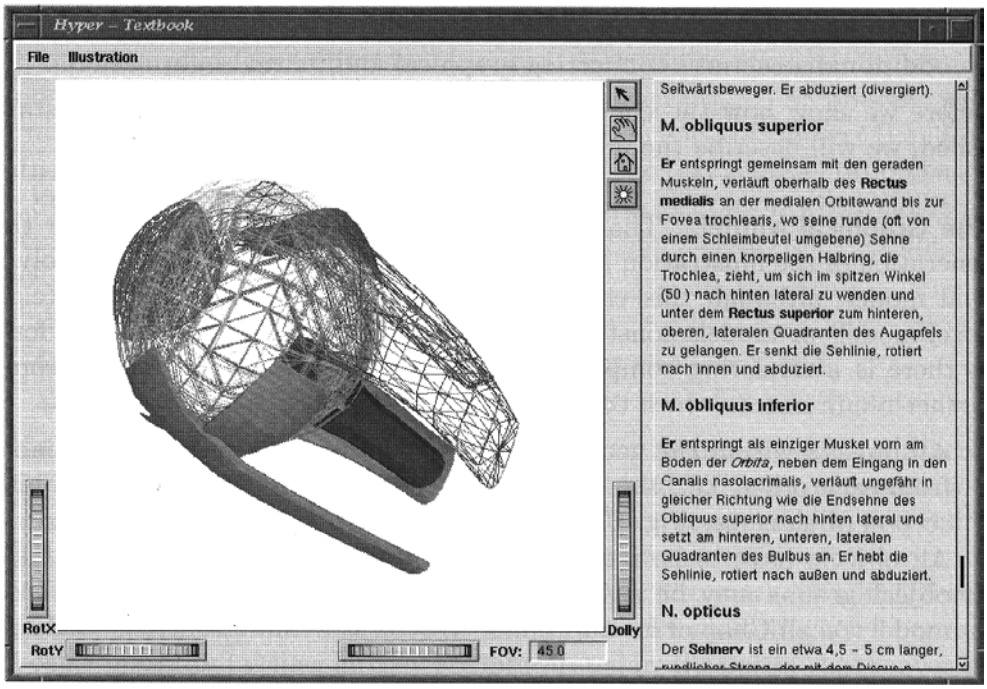


Figure 75: Screenshot of the *TextIllustrator*, a Text-Driven Interactive Textbook on Anatomy

Note that the texts we deal with here are really meaning postulates. They determine, for example, the *concept* of a certain muscle connected with the eye, not the individual muscle of, e.g., your left eye. Thus, they actually correspond already to (partial) T-Boxes. In *TextIllustrator*, the content of the text is not represented explicitly in a knowledge base. Essentially, there are direct links between the medical terms and corresponding parts of the 3D model of the scene. These links are "registered" during the setup process of the application, which establishes in principle the relation between the concepts of single 3D models and the corresponding concepts of anatomic entities.

Nevertheless, the part of the text visible at a time defines a co-textual conceptual context, which could easily be captured by means of a T-BOX of medical sortal objects. Correspondingly, the geometric models underlying the computer graphics can be thought of as a T-BOX of 3D-objects (geometric Gestalt concepts). Scrolling the text changes the textual T-BOX more or less drastically depending on how much of the older text is still visible. We expect a corresponding change in the graphical T-BOX: the system has to determine which sub-models of the complete 3D model are to be contained (together with their locative relations), which is quite simple with the field-external relations implemented by means of the registered links. They determine exactly the new »picture content«.

4.4.1.3 The Selection Problems: Form

Once the semantic core of a picture is determined, an appropriate form for its presentation is selected – the "how to show" part. For language, this selection problem consists essentially in determining which one of a set of synonymous formulations for the content chosen is to be used in the particular case, and which syntactic schemata are to be applied. For pictures, the analogous selection means deciding about the perspective and frame, the presentation styles, and the lighting.

Figure 76: M^{C}CLOUD on the Function of Naturalism in Comics

Obviously, the camera perspective must be chosen in a way that the »picture referents« (associated with the »picture content« selected) are visible, at all. No object should be completely out of frame or totally hidden behind a larger object. Therefore, the point of view may be neither far away, nor too close. Every object must also remain recognizable (as that kind of sortal object). Thus, unusual points of views are to be avoided if their environment does not induce the correct interpretation: extreme perspective shortenings (anamorphotic presentations), e.g., of a rivet seen along its axis, are very hard to interpret as the correct sortal type of object. Edges or corners of objects positioned at different depths from the camera position lead to problems, too, if they seem to meet from the chosen point of view: the object candidates are merged and a proper recognition is difficult. The viewer must also be able to recognize the type of partially hidden objects (and those cut off by the frame). The environment may induce the correct interpretation: a rivet, for example, remains recognizable if only its head can be seen not in isolation but on the surface of a piece of furniture.[65]

Note that we have assumed so far that the complete 3D-model (i.e., the geometric T-BOX) is already determined when the "how to show" aspects are selected. However, the latter usually has to initiate a backtracking – the "what to show" decisions must be revised if the originally selected content cannot be presented in an adequate manner. This may be the case if no valid point of view can be determined: e.g., some objects are always hidden behind some other objects, or too small in the context. Or every plausible camera position shows some objects from a completely unusual perspective that makes it improbable to recognize the object's type. In such a case, the original »picture content« may be split up, leading, for example, to a sequence of pictures with different perspectives, or an enlargement to be used as a pictorial inlay.

The second form aspect – how to select the presentation styles for a stylistically mixed presentational picture – is at least partially also related to perspective, though in a much more general sense. The presentation styles of a picture often encode the attitude of the sender toward the picture content; changing the style for a part of the picture indicates a different attitude, e.g., importance. Let us come back again for a moment to the example *TextIllustrator*. In the graphic, the objects corresponding to the part of the text currently visible (i.e., important for the viewer[66]) are "highlighted", e.g., by means of

[65] For picture riddles in order to be visually enigmatic, these rules are explicitly broken.

[66] Note that we may consider in this example of an interactive system the user (= viewer) as the secondary sender of the message in a pictorial soliloquy.

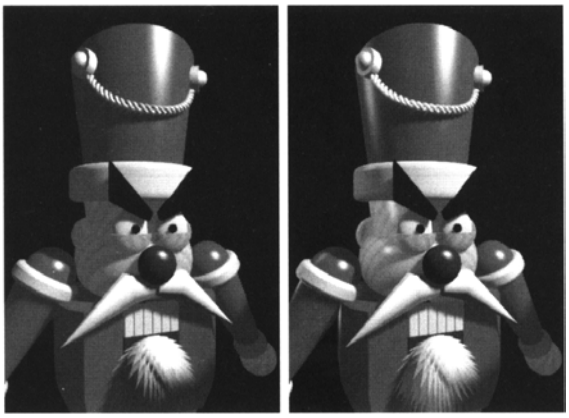


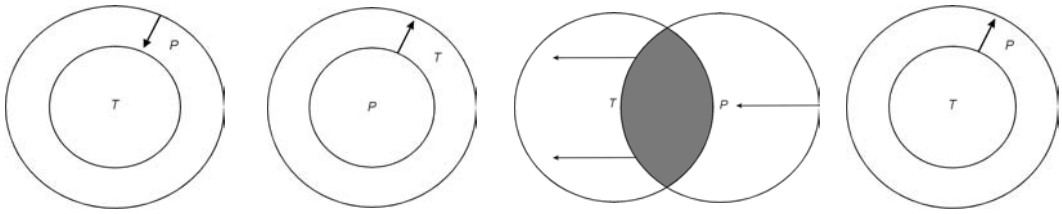
Figure 77: Contrasting Different Lightings – Emphasizing Materiality and Depth on the Right Side

displaying them in color and fully shaded while the others remain gray or use a reduced "technical style" (wire frames, cf. again Fig. 75).

Many comics give a good example of explicitly mixed pictorial styles, and there has been some thinking about the significance of more or less naturalism (in our sense, cf. sect. 3.5.2), as well. S. M^{C}CLOUD [*1993, 44*] associates a more naturalistic depiction of an object contrasting a simplified background with a more objective and distanced view that is necessary if that object as such, not its (regular) use or presence, is focused – if, for example, a goblet is shown as being The Holy Grail. A graphically reduced representation, e.g., outlines only, that does not stand out against the representation style of the background, is usually applied to indicate that there is nothing special about that object, and that it is merely relevant in its normal use – for example, a goblet as a functioning extension for drinking of the protagonist's body (cf. Fig. 76). Similarly, the relative degree of naturalism of a person's pictorial representation can be used to mark that person as either something strange, as somebody in emotional or cognitive distance to the protagonist/reader (more naturalism than background), or as a familiar face and nobody unusually strange (no difference to style of the remaining scene). The protagonist should be drawn in an even more reduced naturalism, compared with the rest, in order to simplify the identification of a comics reader with the character.

This effect of style differences in fact plays with the distinction between medium (context) and figure-ground (proposition) mentioned in chapter 3: while contexts provide the medium of potential figure-ground distinctions, propositions articulate a particular figure-ground distinction on the basis of a given medium. The variation of presentation styles suggests a certain figure-ground distinction over and above the mere medial spatial configuration. While a neutral context builder leaves the figure-ground differentiation to a complementing verbal commentary (or completely to the viewer), pictures with articulate style differences can be read as bearing a preferred reading: take the less detailed or abstracted parts for anchoring purposes; the naturalistic and detailed parts bear the essentially new of the message. Still, the picture does not articulate the figure-ground dichotomy in the unique manner of a proposition; it remains a context builder that induces many more propositions. We may call it a *rhetorically enriched context builder* and come back to this category in the practical context of the second case study in chapter 5.

Selecting the lighting parameters is related to the representation styles. The representation of shadows resulting the lighting is an important clue for depth in the scene and the materiality of the objects (cf. Fig. 77). But they also have subtle attentional and emotional effects. Important parts of the picture are accentuated – "highlighted". This means of rhetoric enrichment was already used in ancient paintings and can be formalized by means of a feedback loop (cf. [HOPPE & LÜDICKE *1998*]). Unfortunately, the



(a) Rule A (b) Rule B (c) Rule A' (d) Rule B'

Figure 78: Differences between Goal Specification *T* and Picture Content *P*, and Rules Associated

emotional effects are far less predictable. They depend highly on circumstantial factors, and hence resist formalization.

4.4.1.4 Combined Selection Problems for Choosing a Picture

For pictures, the two aspects of selecting content and selecting form parameters are less clearly separable than for (propositional) language. A reduced but integral form of the pair of selection problems appears if we only choose among the given entries of a picture database. In this case, form and content obviously cannot provide independent dimensions of free selection: choosing one of the given pictures determines both features at once. Let us assume that we have a set of images each with associated instances of »picture content« and »picture referent«.[67] Content and referents together can be read as a description of what the corresponding picture shows. We furthermore have a similar description of what picture we would like to find. This goal specification may be derived from a text. Let us call the description of a picture *P*, and the description of the specifying text *T*. Comparing *P* and *T* literally and minimizing the differences is certainly a good first choice of strategy. In some cases, we might be lucky: the database contains a picture with a description fitting exactly the pattern – the difference between *P* and *T* disappears. However in most cases, the picture is either too specific, too general, or the two descriptions overlap partially (cf. Fig. 78a to c). What criteria can be brought forward to decide between several of such cases?

L. PINEDA [*2003FC*] deals exactly with this problem in the context of interactive multimedia generation. He extends an idea of VAN DEEMTER [*1998*] who suggests that the picture selection process can be thought of as a deductive inference. Then, the picture investigated could have a more general (i.e., weaker) content that implies the goal specification completely. Or the picture in question has a more specific (i.e., stronger) content that is fully implied by the goal specification. VAN DEEMTER shows that for such cases a strategy can be successfully applied that indeed minimizes the difference from either side. One could use quite effectively the weakest (least informative) picture in the database implying the goal – VAN DEEMTER's *Rule A* (Fig. 78a). Or one could select successfully the strongest (most informative) of the pictures available that is implied by the goal – called *Rule B* (Fig. 78b). In the graphical metaphor of Figure 78 Rule A determines the smallest available *P* including *T* while Rule B goes for the largest *P* included in *T*.

[67] If no individuals that are known already apart from the picture can be associated with the pictorial referents, unspecific referents have to fill that hole (corresponding verbally to indefinite noun phrases with an implicit existential component); note that they still are sortal individuals offering the potential to be found again in other contexts.

PINEDA modifies the two rules for the more complicated overlapping cases. He distinguishes the situations where the communicative focus is on the spatial relations from those emphasizing the conceptual information about objects. Sometimes, it matters most that certain geometrical relations are presented. Even if there are interpretations for the object candidates in the picture that are irrelevant for the communicative purpose, a viewer usually does not consider such interpretations since the identification of objects also depend on his or her expectations in the situation (e.g., induced by co-text). For such a case, a relaxed *Rule A'* leads to good results: chose the picture whose content implies the largest intersection with the goal specification. Figure 78c illustrates this rule.

In other situations, rich conceptual information about objects is to be expressed. However, corresponding pictures can only be used to illustrate very specific situations, i.e., it is highly unlikely to find a fitting image in the database. PINEDA suggests for that case to rather use a schematic picture and complement it with explaining text. For this purpose he suggests a variant of *Rule B* leading to the weakest picture whose representation is implied by the text such that figural and reference objects in locative expressions can be bound to schematic representations of spatial objects in the picture. The arrow in Figure 78d points accordingly from *T* to *P*, since the goal specification is adapted to the pictures available.

4.4.2 Anticipating the Unknown Beholders

The approaches to the selection problems described so far are still mainly semantic: beside a few hints to "the communicative situation" or "the communicative purpose", senders and receivers of the pictorial (or multi-modal) messages have not been considered explicitly. But selecting content and form of a (pictorial) message is always relative to a communicative purpose that is part of the current language game (in the broad sense of "language"). The basic idea for considering the (more ore less unknown) beholders and their understanding of a pictorial sign act planned, is to anticipate that understanding and its presuppositions, and to integrate those anticipations in the generation process.

4.4.2.1 Remarks on the Purposes of Picture Uses

In a formal manner, the purpose of each utterance is determined by the language game in which the utterance is a move – a speech act. The expressions 'language game', 'way of living' [WITTGENSTEIN *1953*], and 'speech act' [AUSTIN *1962,* SEARLE *1969*] (or perhaps less common but more general 'sign act') have become crucial tools for understanding modern pragmatics. Although primarily developed for language in the narrow, i.e., verbal, sense, they span an interpretative schema for any complex sign behavior. Like in a game of chess, languages form rule-based systems of moves: in such a language game, not only

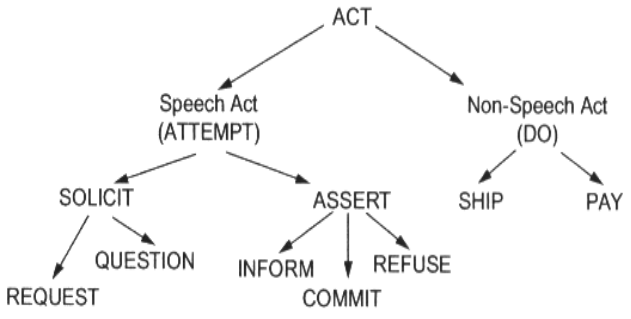


Figure 79: System of Acts to Be Considered in Business Communities

an act of sign use is considered a move, but also the other activities of the participants related to the signs, e.g., applying the sensory-motor test routines associated with concepts (cf. Sect. 3.4.4), a certain coordinated movement in the context of a hunting group, and the ordering of goods between firms (cf. Fig. 79, [PARUNAK *1996*]). The rules in chess determine essentially which moves may follow a certain sequence of preceding moves, and distinguish those regular moves from the illegitimate ones. Perfoming one of the latter interrupts the game.

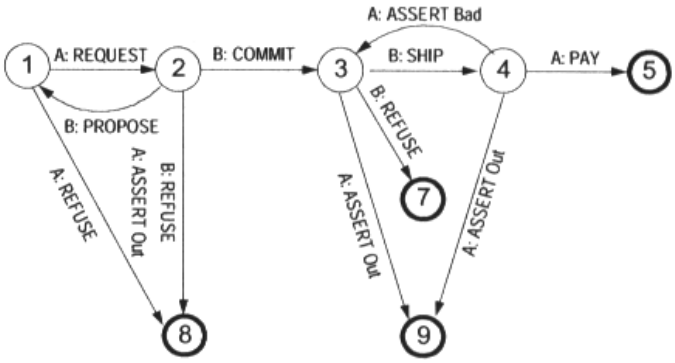


Figure 80: State-Transition Network for Speech Acts

Similarly, signs acts and certain other acts are woven into a structure of regular sequences, while other sequences do not correspond to the way of living determined by that language game. Retaliations have to be expected if a move not sanctioned by the rules is applied. More or less formal systems of rules for the sequences of verbal speech acts can be designed (e.g., by means of Finite State Machines, cf. Fig. 80, [WINOGRAD & FLORES *1986*]), and can be composed hierarchically to form more complex language games. Such formalizations have been adapted for text generating computer systems (e.g., [COHEN *1978*]).[68]

Many, though not all, speech acts have propositional cores. Beside the situational context of the utterances, the propositions are indeed the "glue" binding together the speech acts of a language game in a common discourse universe. The same propositional content may partake in different "illocutionary" functions: as an assertion, a question, a demand etc. Typical speech acts without a propositional core are greetings or excuses, but also the one-word utterances of infants, and warning cries – both of which we have met already as examples of quasi-predicates in Chapter 3. It is usually assumed that many of the illocutionary roles of verbal acts that need a propositional core can also be completed by images instead. Correspondingly, the role of pictures in pictorial sign acts has been compared to propositions, predications or nominations. The act of context-building with its close relations to quasi-predicates, i.e., to non-propositional illocutionary functions (cf. sections 3.4.3 & 3.4.5), has not been considered systematically, so far. Accordingly, current computational approaches to pictorial communication deal mostly with the nominatoric or propositional aspects derivable from context-building.

There have been several approaches to formalize language games for pictorial communication. A transfer of SEARLE's speech act theory to pictorial sign acts was, for example, proposed by KJØRUP [*1978*] leading to a set of conditions for successfully performing several types of "picture acts", among them representation act and illustration

[68] One of the most influential applications of speech act theory in computer science has been the system *Coordinator* by WINOGRAD & FLORES [*1986*] covering action-oriented conversations in organizations. Such a dialog is initiated by partner *A* with a request (resulting in dialogue state 2, cf. Fig. 80). Partner *B* may answer with committing to the request, with proposing a different action or with refusing the comply. The latter results in the same final dialogue state (8) if *A* withdraws the initial request or refuses to the proposal of B.

act. For SACHS-HOMBACH [*2002, Sect. 6.3.4*], illustration (orig.: 'Veranschaulichung') is the most elementary illocutionary act to be performed with pictures, since the primary act of illustration is immediately associated with the picture content, i.e., a (more or less complex) predication. In contrast to the use of predications in propositional speech acts, however, the habit of distinction covered by a concept is not plainly put into the interlocutor's focus of attention: that concept is simultaneously "characterized" by the visual symptoms associated with it. This is, for SACHS-HOMBACH, particularly true for geometric concepts, e.g., the sketch of an ellipsis given in a geometry class. In consequence, the function of illustration becomes less determined if more complex concepts are considered. If an explicit referential component to a sortal entity is to be dealt with, the illustration function becomes the illocutionary role of "exemplification" (orig. German: 'Illustration'). In this case, so SACHS-HOMBACH, the picture does not only illustrate a property but applies that property to an object. The nominatoric identification of the object is usually performed by means of additional texts, e.g. a caption.

So far, state transition diagrams with explicit use of picture acts have not been described: if pictures are used with the same illocutionary forces available for verbal sign acts, they may just be substituted in the corresponding transitions. Speech act transition diagrams are optimal for investigating interactions with a lot of turn-taking. They are less relevant if we are interested in the coherence between the various parts of one complex speech act. While speech act systems focus on the dynamic aspects, rhetoric relations concentrate on the relations between the moves of a language game, and are therefore preferred for analyzing the internal (static) dependencies between the parts of a document. Such relations – for example, "elaboration", "volitional cause", "enablement", "concession" and 19 more in the Rhetorical Structure Theory RST [MANN & THOMPSON *1987*] – correspond to intentions the author looks out for with the utterance relative to the co-text .

Although pictures are originally not considered, they can be embedded in RST by means of the findings of media psychologists (e.g., [LEVIN *ET AL. 1987*]). ANDRÉ & RIST [*1993*], for example, use rhetoric relations derived from linguistics in order to describe the connection between the parts of a picture or between pictures of a series in the context of instruction manuals: a "cause-result" relation often holds between two pixemes of a picture while the relation "elaboration" characterizes the rhetoric link between a pictorial inlay in a main picture in many cases. Some relations relevant for texts are not directly applicable (e.g., "condition", "negation", "concession") but can be integrated by means of conventional pictorial symbols (e.g., arrows, red crosslines). On the other hand, some relations – in particular between pixemes and text fragments like "label" – have usually not yet been considered in linguistic theory. RIST & ANDRÉ organize the rhetoric functions ascribed to pixemes hierarchically according to their abstractness: the top node "depicting an object" is refined to "show characteristic form", "show relative dimension", "show parts", and "show material". Other top nodes mentioned are "show location of an object", "show object state", "show object trajectory", and "show action".

Those rhetoric relations are used as one part of a tripartite proposal to structure the complex sign acts in a multi-modal system. The *rhetoric structure* is basically a tree of rhetoric relations between the parts of the complex sign act. An additional *intentional structure* covers explicitly goals of the sender associated with each element of the rhetoric hierarchy. The sign act hierarchy and the goal hierarchy are usually quite similar but have not necessarily the exactly same structure. For example, subordinate sign acts need not be subordinate goals, as well. Finally, an *attentional structure* is defined by "focus

spaces" each of which is associated with an entry of the intentional structure and contains the discourse entities relevant for that goal – the discourse context in its nominatoric aspect, that is. It is essentially useful for identifying co-referential items.

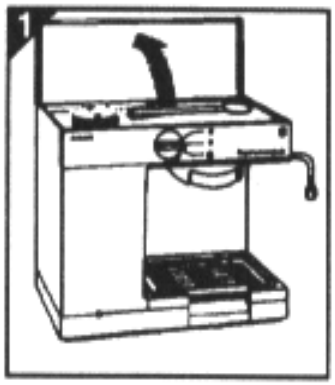


Figure 81: Example „Lift the lid" ("Klappen Sie den Deckel nach oben")

As an example, ANDRÉ & RIST analyze the picture-text combination used in the instruction manual of an espresso machine. The manual in question consists of sequences of combinations of speech acts (in the close sense) and pictorial acts with representational and simple structural pixemes. The first element of one of those sequences ('Filling the water container') is given in Figure 81. The corresponding rhetoric structure, shown in Figure 82, links the three main pixemes of the picture as an auxiliary sign act ("NH") to the text as the main sign act ("HH") of the complex. The function of the picture is analyzed as enabling the user to perform the request articulated by means of 'Lift the lid'. A special sub function is ascribed to the main pixemes: the nomination used in the text – 'the lid' – reappears co-referentially as the result part of a complex relation "Inform-Cause-Result". The arrow pixeme, an abstract graphical symbol representing conventionally an action, takes the corresponding place of the cause. Most interesting for our approach of context building is of course the elementary function "provide background" associated with the third main pixeme. This function is indeed part of every analysis given by RIST & ANDRÉ, the manifestation of a separation of figure and ground already fixed.

The intentional structure (Fig. 83) is closely associated with the rhetoric structure and corresponds approximately to SEARLE's preconditions for corresponding speech acts; each of its nodes describes the (sub) goal associated with a certain partial sign act. In the example, the main goal is analyzed as making the receivers ("users") open the lid. To that purpose the receivers must be able to do so, and also believe that the sender wants them to perform that action. That is, the sub goals are either making the receivers believe some propositions or enable them to perform some action.

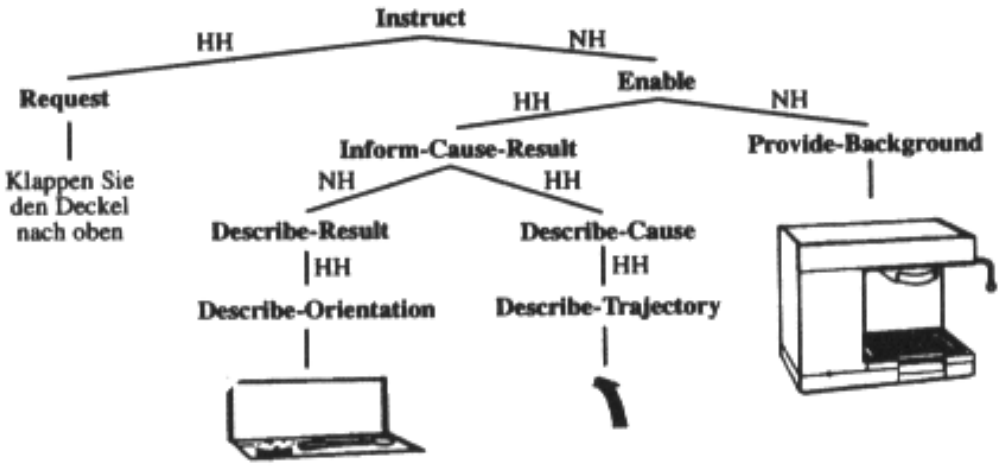


Figure 82: The Analysis of ANDRÉ & RIST: the rhetoric structure for the example (Fig. 81)

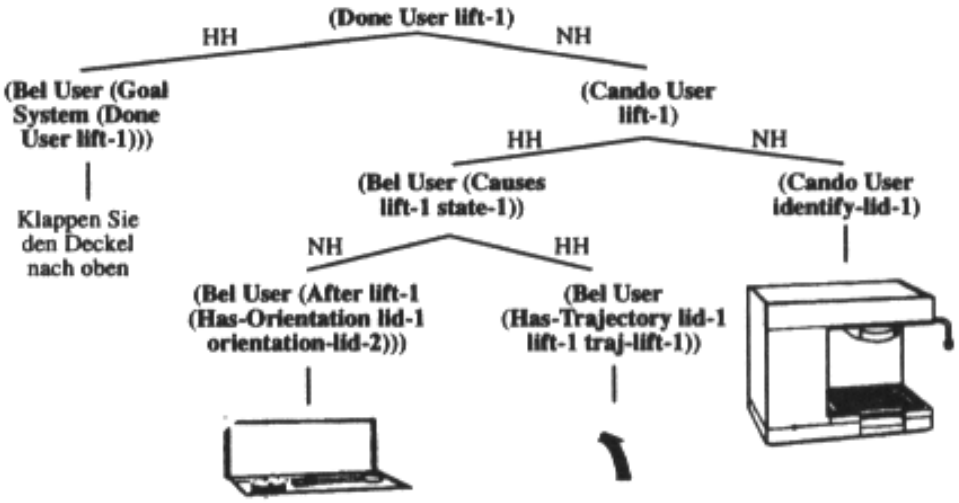


Figure 83: The Analysis of ANDRÉ & RIST: the Intentional Structure for the Example (Fig. 81)

Those analyses are used by ANDRÉ & RIST to specify and build a paradigmatic planning system for the content and form of interactive multimodal presentations [ANDRÉ *1995*], and in particular of the graphics used there [RIST *1996*]. The description of the picture's rhetoric and intentional structures are incrementally derived from a given communicative goal together with further constraints for the generation (e.g., restricted resources) by means of a set of strategic rules. We shall not go into further detail of those strategies. The essential aspect for us is that the receiver's knowledge, believes, motivations, and abilities to perform some actions (as viewed from the sender's perspective) are taken into account as the crucial factors for all pragmatic decisions.

4.4.2.2 User Modeling for Pictures

It was already mentioned in section 3.5.3 that the anticipation of the other interlocutor is crucial to the conception of conscious communication following MEAD. The communicative partners have to anticipate each other's reactions – or a common meaning of a sign act remains impossible since the sender's perspective differs necessarily from the receiver's point of view. The sender has to take into account the receiver's perspective if she or he is to communicate in more than just a signal language – and the receivers must anticipate the sender's position if they do not simply react to a signal but *understand* the sign act as meaningful.

A classical part of pragmatics in linguistics demonstrates these interdependencies in a particularly clear manner: we usually do not communicate everything explicitly but leave out information we know our interlocutors can infer nevertheless. In the context of his investigations on cooperative communication, GRICE [*1974*] christened those special inferences 'implicatures' and formulated several "maxims" that determine certain aspects of cooperativity: any contribution should be, for example, as informative as possible, but also not more informative than is required. Furthermore, the receiver should not be able to infer more from the utterance in its context than the speaker wants him to know (and infer).

With respect to propositional language, implicatures arise essentially from applying syllogistically the meaning postulates associated with the predications to the corresponding contexts and thus adding additional propositions (cf. again Sect. 4.3.1.3):

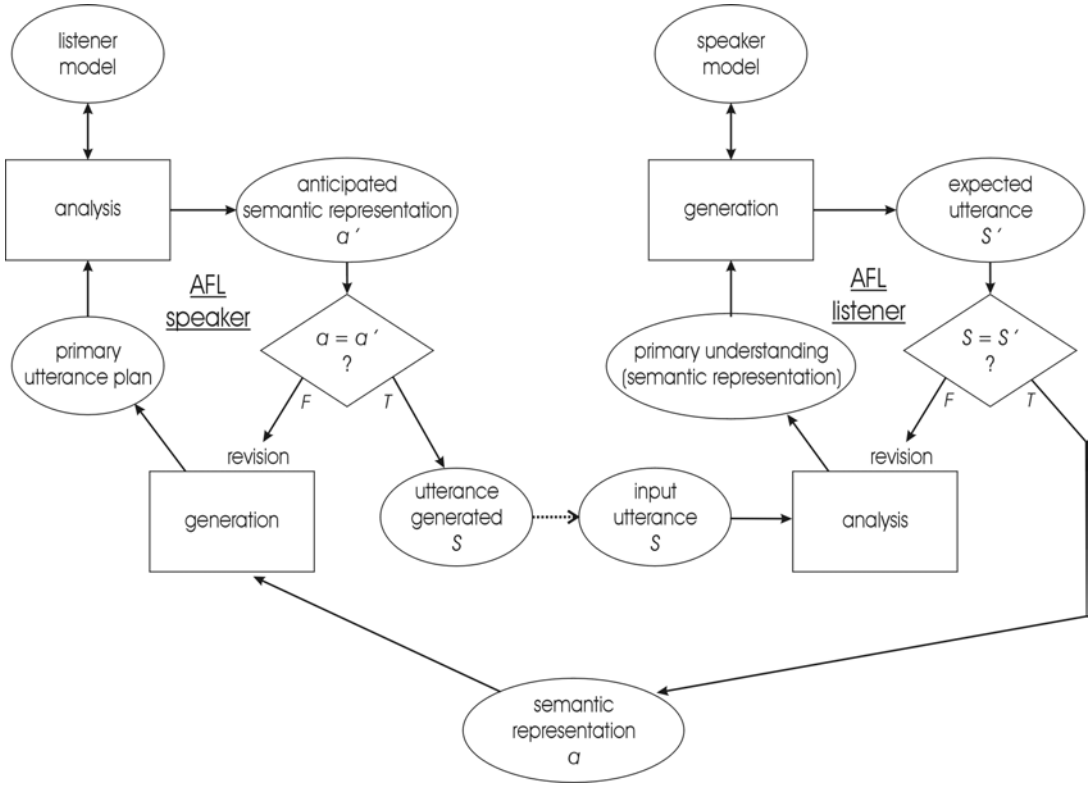


Figure 84: WAHLSTER's Anticipation-Feedback-Loop for Natural Language Systems

if the new proposition states that Buckbeak, an individual in the context in question, is a hippogryph, and a meaning postulate connects the concept »hippogryph« with the concept »being pride«, we may also add the proposition "Buckbeak is pride" to the context.

From a more technical, information-theoretic perspective, STROTHOTTE & STROTHOTTE [*1997, 85ff.*] describe the difference between the information explicitly stated by the sender and the implicatures (expected to be) drawn by the receivers by means of the opposition of (explicitly) *transmitted* vs. (implicitly) *transputed* information. The purpose of the distinction is to conceive of the transputed information as a proper part of the communicated message, as well, in contrast to something additionally generated purely on the recipient's side. That is, the sender is also responsible for that part of the message.

The common approach in computational linguistics is a "user model" anticipating the interlocutors with their reasoning abilities: "What would I understand in his position if being told what I plan to tell him?", and "What might I in her position have tried to tell me with such a message?" respectively. Such a user model works essentially as a "generate and test" feedback cycle. For natural language systems, WAHLSTER [*1991*] has proposed the double anticipation feedback loop (Fig. 84): starting from a semantic representation of what is to be said, the sender system generates a first proposal for a corresponding utterance, which then is analyzed with respect to the meaning a potential interlocutor would ascribe to that utterance together with its implicatures in the context of the previous interactions. After comparing the result of that analysis with the initial semantic representation, the program decides whether the utterance under investigation leads to a sufficiently satisfying communication and is performed, or whether the plan has to be revised first. Indeed, the understanding (anticipated) does not need to cover the

complete initial semantic representation since subsequent utterances could be used to add missing parts or repair misconceptions up to a certain degree.

Similarly, the receiver system generates an initial understanding of an "incoming" utterance in the current context (by the same algorithm of analysis), which then is used as the starting point for re-generating an utterance. Here, the comparison with the real utterance leads to further hints for the interpretation, in particular concerning presuppositions, i.e., additional propositions that one has to assume as being true in the context of the utterance in order to understand the utterance at all.[69] A revision of the initial understanding could be rated as necessary.

Revising the primary interpretation is typically the case for metonymic expressions, too, when, for example, a nurse tells her colleague "our kidney cyst has visitors". Here, a natural language generation system estimates by means of its receiver model whether the receiver intended is able to interpret the metonymic reduction from the full form "our single in-patient who is attended for his kidney cyst", while a corresponding automatic receiver system starts with the noun phrase's literal interpretation as a particular pathology of a (human) organ, finding that the predication is incompatible with that interpretation, and thus activating the revision of the interpretation as a patient who is able to get visitors and who the speaker could have metonymically associated with the literal interpretation.

Under the precondition that a propositional analysis of an image to its picture content (and the derived referents) is sufficient, the general schema of linguistic user modeling can be transferred to pictorial communication and in particular to tele-rendering with only slight alterations: the active and passive beholder models (Fig. 85). After generating a picture from a primarily selected picture content (e.g., by means of a strategy-based planning algorithm like the one of ANDRÉ & RIST), the sender system uses its model of the passive beholder to analyze the picture by means of algorithms corresponding to those described in section 4.3. The analysis results in the »picture content« that the intended beholders are presumably able to decode when being shown that picture: their anticipated understanding. Furthermore, referents are derived from the content either as purely intentional individuals or as individuals that are known from other contexts and fit the description in question. A comparison between the two instances of »picture content« corresponds closely to the one between the semantic representations in the verbal case: two sets of (spatial) propositions – either literally "given" in the picture, inferred as implicatures or induced as presuppositions – are compared with each other. The differences are used to revise the picture initially generated or to reject it altogether and start up all over again. If, for example, a referent is not recognized because the passive beholder model does not dispose of the necessary relation between sortal object and geometric Gestalt in question, the plan has to be revised so that the beholder (model) becomes able to recognize that geometric projection as one of this referent, e.g., by deriving the unknown perspective from a known one in a preparatory picture sequence.

On the other hand: having been presented a picture, the (passive) beholder reaches a set of spatial concepts (together with their geometric test routines) as the primary »pic-

[69] For example, the proposition "The president is not intoxicated today" (uttered in a context so far thematically neutral with respect to intoxication) induces the presupposition that the president in question is usually – or at least often – drunk; the utterance would not be informative otherwise since the default assumption in the contexts we have in mind here is that presidents are usually not drunk. If used nevertheless, the receivers must conclude that the sender intended the presupposition, as well.

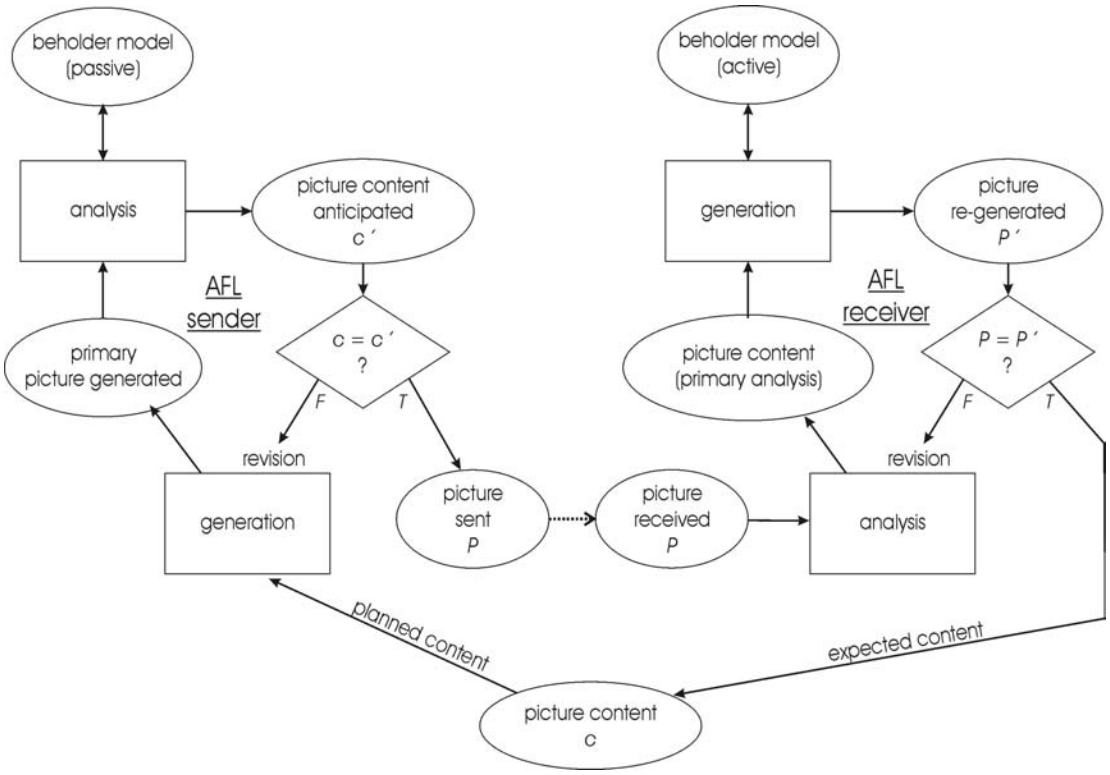


Figure 85: The Two Beholder Models for Tele-Rendering

ture content«: by analyzing the picture on the basis of clustering geometric Gestalts (bottom up), and by projecting sortal object models (top down). Those concepts may be applicable to some individual sortal objects as the picture's referents known from another context. A generic intentional object of the appropriate type may be generated as a default referent instead. The set of propositions thus formed is extended by means of inferences and inductions. In order to render the inferences to implicatures, and to transform the inductions into presuppositions, the receiver system's model of the active beholder takes the expanded picture content and generates – from the assumed sender's perspective with the potential intentions – another picture: the image the receiver actually expects from the assumed sender in this communicative situation. The differences between the two pictures indicate inconsistencies in the primary interpretation or in the sender's intentions or background postulated by the model.

There arises one complication at this place: how to compare the two pictures. "Syntactically, e.g., by pixels" may be the immediate answer – an answer obviously without much value: on the level of pixels, a minor difference could mean a completely different picture, and a major difference nothing more than a slight darkening of the whole image. High-level pixemes are certainly much more informative; but remember that such pixemes are only accessible in combination with semantic categories.

The reasonable solution would be to add another complete analyzing step, and then find the differences in the corresponding instances of »picture content«.[70] The full form of simulative beholder modeling for graphics in media of class *IV* is in fact prohibitively expensive: after all, picture analysis and picture generation each demand already for quite extensive computational resources. Correspondingly, the combination of both in a potentially iterative feedback loop has never been realized so far. RIST, for example, by-

[70] We come back to the realization of such an iteration in the case study of Section 5.4.

passes the analysis of a picture to be generated by the sender's passive beholder model. While incrementally constructing the plan for generating the image, i.e., while determining successively parts of the picture's content from the sender's point of view as a description in propositional form, a secondary description is constructed in a rule-following manner covering the picture's content from the receiver's perspective. Note that this second description is not generated from the picture but mirrors, so to speak, the underlying reasoning processes of the sender with respect to a potential receiver. For each step in its design activities, a rule tells the system how the corresponding change in the resulting picture will (presumably) change the understanding of the beholder in question. The picture *per se* is not at all involved; no geometric test routine has been employed.

Quite obviously, there are some further complications with explicit beholder modeling, which should at least be mentioned at this place:

- Variations in the type of beholders anticipated: in a positive sense, beholder models for tele-rendering are, thus, adaptable to different user groups, and can react to lay persons, novices, experts in the domain of discourse. In a negative sense, an incompatible or wrong user model impairs the performance severely.
- Iterations of the revisions: revising the initially generated picture (or picture understanding respectively) usually improves the pragmatic quality of the product. But there is no guaranty that the revised picture is sufficiently understandable in the correct way, or that the interpretation reached can be trusted without hesitation. Quite in contrary, the revisions have to be tested as well. So: How many cycles are necessary? Does the revision always improve the performance so that the approximations come close enough after some steps? Or is the improvement not monotone, and the successive revisions jump in cyclic (or even chaotic) alternations? There is no general answer known to these questions so far. In individual cases, monotony can be ascertained.
- Recursion of modeling: if a sender models generally a receiver, and *vice versa*, then, this sender model should include again a receiver model, which should include a sender model, into infinity (and *vice versa*). The sender believes something about the receiver, and he believes that the receiver believes something about the sender, etc. An explicit simulation of those recursions is not possible. Theoretically, the conjunct of this infinite chain together with its complement is summarized in the user modeling community by the predicate 'mutual belief'.

These complications have been dealt with extensively in the literature on user modeling. Since they are not specific to the uses of pictures, a more detailed description is not necessary in our context.

Any further processing of high-level illocutionary aspects of picture uses, e.g., the symbolic representation of peace as Noah's dove, can then indeed be viewed as fully equivalent to that of verbal signs. The only difference is the initial transformation to the propositional form of combined picture content and associated pictorial reference. Those illocutionary aspects are, thus, not specific to pictorial sign acts and of limited relevance for our investigation, as well.

This kind of beholder modeling depends on whether pictures are really just determined by their content (and the reference relations derived from that content). Are pictures indeed completely covered by means of a definite set of propositions? Or is the content itself a product of a much more fundamental communicative function characteristic for pictorial sign acts?

4.4.2.3 Adaptation to the Pragmatics of Context-Building

The last section has sketched the treatment of pictorial pragmatics in computer science in its present state: the approaches are essentially adapted to the structures of verbal assertions for which solutions have been developed before – independent from pictorial communication. And indeed: although the nominatoric aspect is far less clearly separated in pictorial sign acts compared to an assertion, the principal distinction between substance concept and attribute concepts of the underlying sortal field often induces a first interpretation of representational images as a more or less complex assertion predicating properties of and relations between (i.e., attribute concepts of) the instances of sortal objects (the substances of the field). In rhetorically enriched pictures, form aspects are furthermore employed to modify or focus the nominator/predicator structure determining the interpretation of the picture. Therefore, an adaptation of pictorial sign acts to verbal assertions in user modeling seems rather reasonable at first, and in fact leads to satisfying results for special tasks.

In the light of the discussion outlined in chapter 3, the presentation of a picture constitutes in general a communicative act with some characteristics quite different from the utterance of an assertive sentence. If perceptoid signs are conceived of as a special kind of context builder, their function does not yet include a particular figure-ground distinction or a certain differentiation in nominatoric or predicative aspects. The beholders – and possibly some accompanying sign acts, in particular by means of assertions that take that picture as context builder – induce both. Metaphorically speaking, context builders provide originally the medium in which various figure-ground distinctions can take place. It is, thus, crucial for pictures that they are in principle open to various interpretations – evoked by subsequent sign acts, be they performed externally or as an inner soliloquy. An adequate user modeling for pictorial communicative acts in general must account for this feature, especially if it is intended – like our generic data structure – as a reference model, of which concrete systems realize selected parts only.

Recall at this place the four modes of reflection associated with pictures mentioned in section 3.4.1: the immersive mode, covering standard pictorial communication, is constituted as a specific combination of the more elementary deceptive and symbolic modes. The intuition that pictures – at least the representational ones – do resemble a situational context, i.e., evoke erroneous impulses in our immediate behavior, is the basis of the deceptive mode. It is still »resemblance$_{\alpha}$« we are considering here: the detectors associated with some (pre-) object concepts have to be activated spontaneously by the syntactic properties of the picture (recall the birds of ZEUXIS). Only in the combination with the symbolic mode, i.e., the beholder's awareness of taking part in a communicative act in one role or the other (and indeed, following MEAD, both roles simultaneously), the concept »resemblance$_{\alpha}$« is transformed to »resemblance$_{\beta}$«: the awareness that the activation of the detectors has happened in the wrong situation, and the use of this awareness to focus the awareness of the communicative partners to a different situation in which the detectors are not falsely activated. Only in this combination of deceptive mode and symbolic mode into the immersive mode, the reactions, which are still fixedly bound to the detectors for cases of pure »resemblance$_{\alpha}$«, can be suspended from immediate performance without blending them out of the behavior completely.

Exactly these spontaneous but (more or less) suspended reactions, which also include emotional aspects beside the cognitive ones, enable us to employ pictures as primary context builders; they mark the essential difference between propositional and pictorial communication; and they lead directly and simultaneously to the most acclaimed virtue

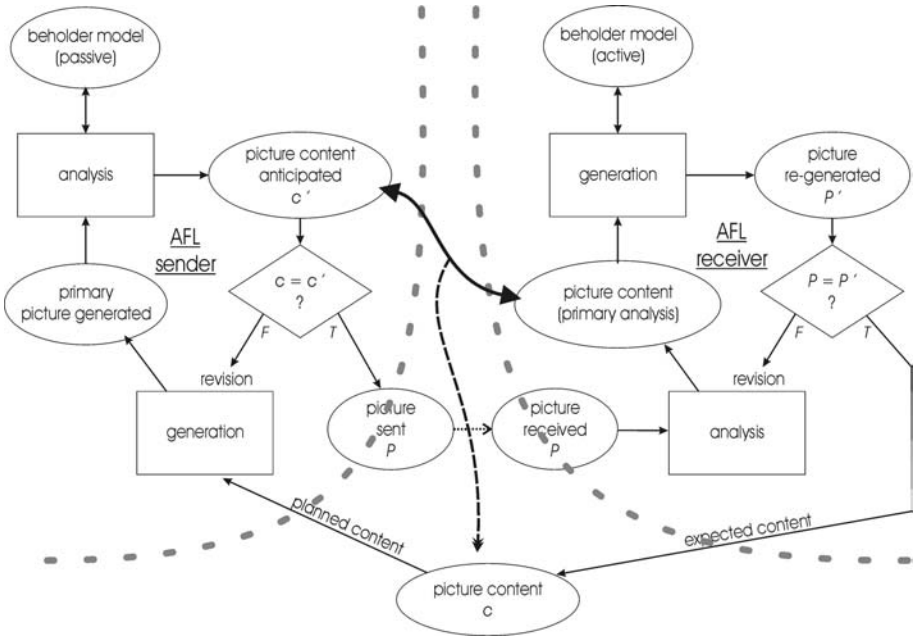


Figure 86: An Objective Picture Content Depends on the Dyad of Beholder Models

and the most decried vice of picture use: the efficiency to communicate complex affairs holistically, and the power to undermine the rational structures of discourse. Complaining about the latter would indeed not make much sense if pictorial communication were essentially propositional. It is, however, quite reasonable with respect to perceptoid context builders, which do not offer a fixed figure-ground distinction but the potential of many spontaneous reactions "from the stomach".

No doubt, a *full* simulative beholder model has to provide a comparable complexity of interacting reception modes. The symbolic mode is essentially caught by the recursions of beholder modeling. For this mode indeed, the pure propositional description of the picture together with propositions covering the beholder's knowledge and intentions (up to the mutual beliefs) is sufficient and provides the distance from immediate reactions on the affairs asserted, that is in particular the picture's predicative aspects covered by the content.

We have to keep in mind that for context builders, content can only be momentarily assigned. It must remain revisable. In the case of ambiguous figures, for example**,** the content changes frequently, spontaneously, and in a rather dramatic manner. This corresponds to the observation that pictorial sign acts do not have *a* propositional content at all – they usually have many of them. As context builders, pictures are not equivalent to a single assertion or any of its (traditional) parts for nomination or predication. There is a much closer similarity to complete texts, like a novel viewed as a rather complex context builder for subsequent utterances. The possible reactions to the presentation of those two types of context builders are at least partially comparable. For example, a fixed linearization to a subsequent re-narration is not included in either case, although the linear composition of the novel induces a certain sequence of assertions much stronger than any picture would do. In general, many paths are possible: in this sense, all context builders offer narrative or argumentative maps. Reducing a picture to a sequence of propositions, as complete as it may be, does not take full account of the pictorial sign act's semiotic potential. It is indeed crucial to not cut off the connection to the

perceptual processes generating the propositional form since they are necessary for applying further test routines.

Employing the same algorithms of computer vision in the beholder model that are used for analyzing directly a scene is certainly the basis for the model's deceptive mode.[71] More precisely: those algorithms mediate an association of a picture content and the picture in question – an association that we view as a description of the anticipated deceptive mode of a beholder. After all: the beholder model always contains descriptions only of the simulated events or states.

Quite obviously, the spontaneous reactions we would expect being named here have to be, in a way, hidden in the instance of »picture content«: let us assume that, for example, adding the item »snake« to the beholder model's »picture content« is rather interpreted as the activation of the *detector* »snake« together with the relevant repulsive movement and other emotional reactions. The execution of such reactions would have been automatically triggered "outside of the model", since they correspond to the intentional pre-objects perceived in the situational context. The entities contained in this variant of »picture content« are transformed to real concepts only by means of their embedding in the beholder model with its constituting role for the symbolic mode. Indeed, we now see that the "free" version of »picture content« described in the section on semantics, i.e., the one not bound within the mutual modelings of the beholders, is not at all sufficient for pictorial semantics, since it cannot really account for the complicated interaction of deceptive and symbolic modes inherent in picture uses. Even the primary »picture content« of the picture producer describing the spontaneous reactions is symbolic only in as far as it is part of the sender's soliloquial communication, and thus stands in relations with the models anticipating the reactions of potential beholders.

The conception of a singular »picture content« developed in section 4.3 merely covers the perceptoid aspects of the concepts governing a picture's semantics. It does not cover the full range of communicative functions usually associated to the concept »concept«. Those aspects are dealt with only by means of the interdependent dyad of »picture contents« in the active and passive beholder models (Fig. 86).

This does, of course, not mean that the simple, semanticist version of »picture content« is not very useful and even mostly sufficient for many practical applications of computer vision where an explicit reflection of the complicated interplay of reception modes of pictorial communication is not necessary for solving the tasks at hand. Similarly, the intralexical definition of verbal semantics is deficient compared to reference semantics; nevertheless, it can be employed with good results to deal with certain problems of computational linguistics.

4.4.3 Authenticity and Media of Class *IV*

Let us take up again the problem of Section 4.4.1.1: an essential aspect for media of class IV related to partner modeling must obviously be the question of authenticity. To what degree can the users of an interactive system trust that the sign act they think they are performing with a picture presented by the system is uttered authentically?

The discussion on authenticity in computer science is usually more restricted in its focus since the term 'authenticity' in the technical sense refers essentially to the question whether the apparent sender of a message is the real sender. Traditionally, a message is encoded in a way that marks it uniquely as from a certain sender – the signature

[71] This holds at least approximately – though, recall the remarks in section 4.3.2.3.

or the seal of a letter, the secret code of an encrypted text, or an individual stamp on a picture. A short introduction on special syntactical changes to images known as 'digital watermarks' [DITTMANN *2000*] reflects a well-known approach of computer scientists to the problem of technical authenticity of images in the second part of this section.

4.4.3.1 Beholder Models and Authenticity

Recall the determination of authenticity in sections 3.5.2 and 3.5.3: a sign act is called authentical if the attitude of the sender to which the sign act shifts the attention of the receiver is the actual attitude of the sender. A proposition, for example, is authentical, if the sender believes that it is true and is ready to argue for its truth. While the category of truth cannot be applied directly to picture acts, authenticity can. To that purpose, we obviously have to know who counts as the sender of the pictorial sign act. We therefore have to investigate whether or in what respect partner modeling is able of helping to improve the receiver's rating of authenticity of tele-rendered pictures.

There are two typical classes of senders in the case of a tele-rendered picture: the computational visualist originally designing the interactive system in question and thus being responsible for the pictures potentially shown; and the user employing the interactive system in a complex soliloquy, hence responsible for the pictures actually generated. Authenticity in the general, communicative sense indicates a specific relation of correspondence between the sender's sign activities and her or his other behaviors. The sender's attitude to which the sign act primarily refers (and which shows in his behavior – his moves in the sign game) must be his or her real attitude, or the sign act is not authentical. In the case of a proposition, it may still be true, even if the sender does not believe that. Since pictures are neither true nor false, authenticity remains as an essential criterion for successful pictorial communication, in particular when using pictures that have been generated by tele-rendering.

Obviously, an interactive program cannot inspect the sender's actual attitude – it is not given the opportunity to observe the system developer's behavior, nor is it capable of reading a user's mind who directs soliloquially his or her thoughts by means of the interactive system's creations. It only has the representation of a small range of potential attitudes in the form of the descriptions of intentions, beliefs, and knowledge in its active beholder model. Accordingly, interactive systems are not able to establish authenticity of sign acts by themselves. The active speaker model can only help to set up well-known standard situations of communication: then, authenticity can be assumed in a generic way not bound to an individual sender, and indeed independent from the two distinct classes of senders to be considered. Again it becomes clear that communicative aspects of picture communication are adequately handled only by means of the embedding in the mutually associated beholder models.

An explicit conception of beholder modeling as described in the last few pages has never been realized so far. From the engineering point of view, reduced derivations of such an extensive modeling as described before are often useful and functional to a satisfying degree for specific tasks, and more efficient, too. Nevertheless, discussing the full complexity pictorial pragmatics impresses on beholder models is obviously important for computational visualistics from the perspective of structural science.

Figure 87: From Left to Right: Original, Watermarked Original, and Watermark Image Used

4.4.3.2 Authenticity as a Technical Problem: Syntactic Approaches

When the expression 'authenticity' is currently used in computer science, it does not refer to the coordination between a sender's attitude and the message's content; that relation is usually not accessible for the programs. There are, however, commitments of the computer as a medium,[72] among them those called 'integrity' and 'authenticity'. Integrity is granted if the receiver of a message gets exactly what the sender has sent, i.e., nothing has been left out, added or changed. Authenticity in the technical sense means that the receivers can be sure that the sender marked on the message is indeed the one who has sent it. Signatures are a common means of authentification (e.g., of letters, works of art). In combination, the two commitments also garantee that the sender cannot deny to have sent the message in question (i.e., 'non repudiation' of the message).

Well-known examples of using an explicit authenticity marker are the icons of TV senders superimposed in a corner of the images transmitted. Copyright is another important application area for techniques to ensure authenticity. Since notes of money are particularly well protected against un-authorized copying, many techniques for ensuring the authenticity of a note are exemplarily applied: the use of special materials, additional information in the form of serial numbers, the signatures of few authorized persons, and in particular the means of watermarks.[73] The latter gave the idea for integrating more or less obviously information about the sender's identity in computerized images (and other kinds of computerized perceptoid signs) called 'digital watermark'. For pictures, the watermark is another picture, mostly a binary one, i.e., in black and white.[74]

With a few exceptions, digital watermarks should not be obvious (cf. Fig. 87). Not being perceptible for someone who tries to illegally copy or manipulate the message is a first step of avoiding the watermark itself from being manipulated or removed. Relative to the tasks the digital watermarks have to fulfill, several types can be distinguished depending on who should be able to detect the watermark and how robust it must be. A watermark is robust if it is as complicated as possible to manipulate or remove the additional information embedded as watermark in the image from the message (without destroying the message). This includes "friendly" manipulations like compression or encryptions used for transmission.

[72] More precisely, those commitments are assumed by those providing the medium.
[73] Although exchanging money is usually not conceived of as communication, money has quite obviously very much in common with a sign.
[74] The watermark image may or may not show text.

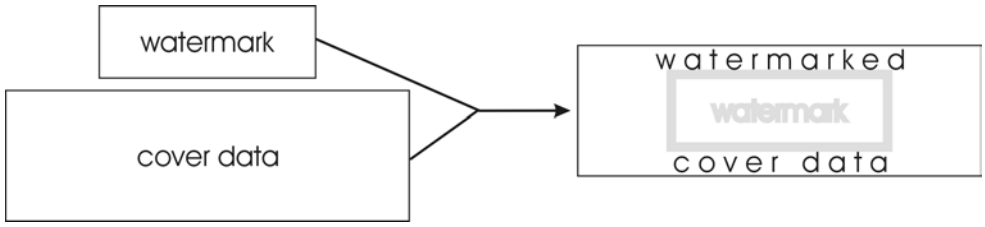


Figure 88: The Principle of Digital Watermarking

If the purpose is mainly to inhibit unauthorized copying, the digital watermark may be mainly detectable for the sender only. Of course, it must be robust in case the copiers have noticed the watermark.

If watermarks are used to ensure authenticity in the close sense they should be visible for the receiver only who wants to detect if a fake message was delivered. Again, they have to resist attempts of manipulation or removal, but must also be hard to fake. Therefore, usually encrypted information is used to which only the receivers have the key.

If integrity is the goal, the digital watermark has the purpose of indicating to the receiver any manipulations of the original message;[75] it, thus, should not be easily visible for anybody else. In contrast to the cases above, it may be highly sensitive to "unfriendly" manipulations since any changes the receiver detects in the watermark indicates a manipulation of the message not authorized by the sender. Advanced systems of fragile watermarks even allow the user to identify the regions that have been tampered with. Quite obviously, the watermarks should not be too sensitive: if, for example, the image was encrypted or compressed (without loss!) during transmission, we would expect the watermark to be unchanged despite the manipulations in between. Of course, all three tasks often are at hand simultaneously. So, different kinds of watermarks may have to be integrated into the picture's syntax.

Technically, digital watermarks are a kind of steganography. In contrast to cryptography, where it is obvious that an encryption has taken place, steganography (from the Greek words for 'hidden writing') is the general term for hiding a message syntactically within another message – the cover message – so that it remains unclear to anybody intercepting the cover message that there is a hidden message at all (Fig. 88). A typical example often employed in detective stories is to hide the actual message in the first letters of the words of a cover message with a plausible but relatively irrelevant content. While the textual cover message can usually be constructed as needed, the cover is already given in the case of digital watermarks: it is the original image, which should not be modified perceptibly by the hidden message, the watermark data, if the message's function is to be fulfilled. That is, the syntax of the picture must not be changed too much. Since digitized pictures usually contain a remarkable amount of noise, which is barely noticed by the users, steganographic procedures often hide watermark data in image or sound files as noise; that is, they let the watermark appear as innocent stochastic disturbances. Often, the watermark image is even generated from the original (cf. again Fig. 86): then, large segments without much color variation in the original do not have to carry too much structure of the watermark that could be detected more easily in the homogenous area.

[75] Ironically, most techniques of watermarking do indeed manipulate the picture, as well. It must obviously be an important goal to keep the differences originally created by the watermark at low level. Alternatively, the receiver must be able to separate the watermark and completely reconstruct the original.

For example, a technique called LSB (for Least Significant Bit) is used to lightly modify the color information of each pixel. If the watermark is a binary picture, only a single bit of the code for the color value of each pixel has to carry the value of the same pixel in the watermark. If we select the bit in the code that shifts the color merely to an immediately neighboring one, the picture only differs slightly from the original – a manipulation that is almost impossible to detect with bare eyes.[76] Alternatively, a secret algorithm can determine only some of the pixels for carrying watermark information in their LSB. To extract the watermark information, we would simply need to take all the data in the relevant LSBs of the color bytes and re-combine them.

Patchwork algorithms change sets of pairs of algorithmically chosen small segments of the image ('patches') slightly, e.g., by lighting one patch of a pair a bit and darkening the other. If one knows which pairs have been chosen, a statistical evaluation of the patches can already decide whether the picture has been watermarked, even without knowing the original image. Unfortunately, the patchwork watermarks are not very robust and can be destroyed easily so that they cannot serve for integrity markings, too.

In section 4.2.4, Fourier transformation of pictorial syntax has been mentioned. It leads to another "picture" that can be transformed back without loss. The watermark can therefore be applied to the Fourier transform instead of the original image: since local changes in the Fourier transform are equivalent to small changes at every location in the original image, the watermark is invisibly "smeared" across the picture. It only becomes obvious if the Fourier transformation of the picture is analyzed.[77]

A further application of digital watermark technology is hiding additional information about the picture, its content or its history. Assuming that such secretly embedded annotations are not misused for false descriptions, they can be helpful when automatically searching for pictures, e.g., in the Internet. So far, only the file name, which is often not informative, and the text appearing on the same web page are indicators of the picture's content for the search engines. This can also include warnings about certain kinds of content, e.g., pornography, so that filters can be installed for prohibiting access to such image files, for example by minors.

* * *

In the following, pragmatic considerations for the two remaining classes of pictures – structural and reflective images – are discussed. For the latter, the fourth mode of reception of Section 3.4.1 is important: so far, we have not yet considered the role of that reflective mode, as we have called it, for beholder modeling, and finally come back to it in section 4.4.5 after having had a look at the rhetoric's of structural pictures.

[76] Obviously, the original image has to be given in one of the usual image formats with high color resolution. The LSB procedure does not work well with palette-based picture formats since changing even the least significant bit of the palette numbers may dramatically change the color.

[77] Since some compression algorithms (including jpeg) manipulate the Fourier transforms of pictures by high pass filtering, the watermarking should not be done in the high frequencies, or the watermark is not robust for that compression.

4.4.4 Information Visualization and the Rhetorics of Structural Pictures

It has already been mentioned in section 3.5.4 that structural pictures are pictures based on a shift of meaning between two fields of concepts. The rhetorical instruments of metaphor and, in the case of pictorial abstractions, of metonymy have been identified as crucial for the understanding or generating of structural pictures. The metaphorical use of geometric Gestalts for visualizing contexts of a field of concepts different from the sortal field poses several questions that are mainly handled in the domain of information visualization.

We do not deal in the following with the preparation of the data to be visualized: often, "raw" data is the real starting point – that is, instances of a data structure with relatively little internal organization (e.g., just sequences of triples of integers). The first problem for an information visualizer is, then, to organize (= interpret) that data into a high-level field of concepts. In a second step, the instances of that type are projected metaphorically to pictorial syntax (visualized) – the step we are only interested in here. Since the first step is often quite unclear at the beginning, its solutions can be seen as the main scientific advantage gained by the visualization ("visual data mining"). In such a case, the cycle of the two steps is usually iterated. Profitable conceptualizations are generated in an approximative manner. Intermediate results are gained by experimenting with the form of the conceptualization and with the parameters of the visualization: the pictures help the data mining engineer to perceive unusual patterns in the raw data, e.g., dependencies between different properties. They, thus, induce ideas that eventually enhance the conceptualization. Quite obviously, the crucial point in such an explorative use of pictures cannot lay in a straight-forward soliloquial propositional sign act; even the interpretation as a predicative act (pure picture content) is not quite reasonable in this situation since the picture is originally constructed without the a-priori knowledge of the corresponding concept being applicable here – quite impossible for soliloquial sign acts. The communicative function as a perceptoid context builder, however, is highly plausible: a symbolic situative context is provided in which the data mining engineer can use the pre-attentive grouping principles governing visual perception in order to find high-level structures not known before.

Quite obviously, the first step of information visualization is of an enormous complexity. Additionally, it is not directly concerned with pictures. We therefore assume that the data to be visualized is already organized in a relatively complex data structure that has to be metaphorically mapped into a picture.

4.4.4.1 On Source Domains and Target Domains

In the scientific discourse about metaphor, the two fields of concepts involved are specifically called 'source domain' and 'target domain'. In a metaphor, the internal structure of the source domain is borrowed to the other field (so to speak): with it, the internal structure of the target domain is verbally mirrored. Thus, the expressions that are used in a metaphor are originally for speaking about the source domain but are now employed to mention the target domain.[78] The metaphorical binding between two fields of concepts establishes a field-external relationship similar to the constitution relation we have meet in Section 4.3.1.2. There, we found out that this relation enables us to

[78] The naming is emphasized here, because the association of 'source' and 'target' of a metaphor may seem on first view counter-intuitive: one might prefer thinking of a projection of the target domain's structure to the source domain's expressions, which would motivate the inverse naming.

motivate the meaning postulates of a field by introducing them as a combination of other fields. It also allows us to inherit sensory-motor test routines from the constituing fields to the field constituted. Like the constitution relation, the metaphoric binding provides the target domain with additional sensory-motor test routines from the source domain; and it also allows us to substitute reasoning schemata from the source domain for those from the target domain.

By characterizing the source domain, we can distinguish two general types of metaphors for information visualization:

[1] the sortal field is the source domain: the main pixemes of those pictures are derived as for a realistic scene: they correspond to momentary Gestalt projections of sortal objects, possibly together with auxiliary Gestalts that depend directly on the sortal individuals like shadows or clouds (particles). However, each instance of a sortal object contained represents another entity. The geometric properties, the recognizable parts, the spatial relations, and the color values (in the wide sense) are supposed to have a corresponding meaning in the target domain.

[2] the geometric field is the source domain: geometric entities, like circles, rectangles, irregular shapes, and arcs, together with their sizes, positions, orientations, and/or connected neighbors form the primary picture content; a spatial interpretation in the narrow sense (supported by sortal objects) is *not* implied. Each geometric entity stands for an entity of the target domain, while its metric, topological or visual properties encode corresponding properties of the target domain.

A tendency to associate a history to the entities is inherent only for sortal objects. The beholders expect them to have a development from some contexts into some other contexts. However, just a momentary snapshot of that lifeline is given to them with the picture. This has an important consequence if more than one visualization are to be presented. Identity of entities can be implicitly induced by using a sortal object as source. For geometric entities, identification has usually to be explicitly made by means of labels. A geometric source domain is, on the other hand, in many cases easier to interpret since the rather complicated processes of object constitution are not necessary for the understanding of a corresponding visualization.

Additional pictorial abstractions like emphasis of contours can be in reign only in case [1]: aspects of the target domain are projected to the attributes of sortal objects, but not all of the attributes of the latter that lead to visible features in a picture may be part of the metaphoric connection. In order to avoid confusion of the beholders, such attributes should be suppressed. On the other hand, indications of an interpretation as sortal objects – like shadows – distract the beholder when a structural image with geometric source domain is presented. Hence, the conception of a visualization being *expressive* is important: the complete description of the target domain is to be visually presented, and nothing more (no representational artifacts).

Figure 89 exemplifies the same data in three different structural pictures: obviously, the left image uses a sortal source domain (bundles of stakes), while the right example merely employs geometric entities. The analysis of the picture in the middle is a bit more complicated, since the colored fields could be interpreted as three-dimensional clouds with varying density of very small sortal particles in front of a neutral background. The scale at its side however indicates that a purely geometric 2D interpretation is intended. Every pixel, including those of the (blue) “background” and of the (red)

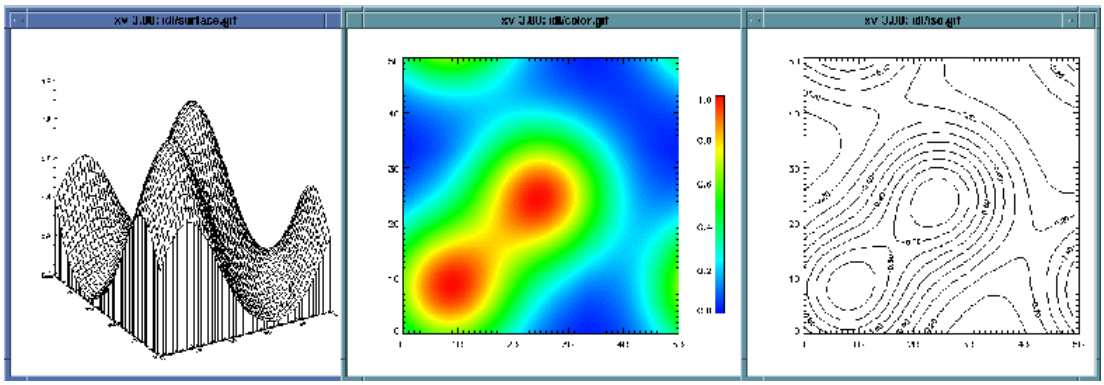


Figure 89: Data of One Target Domain in Three Source Domains

"cloud centers", encodes the same kind of entities (e.g., a location): the color only indicates that they have different values in another attribute.

The target domains can be similarly employed for subdividing visualizations depending on the amount of pictorial syntax they already encompass. The target domain may contain:

(a) sortal objects: the spatial distribution of invisible attributes of sortal objects belongs here where the different attribute values correspond to color values, e.g., the distribution of a house's caloric flow; or radiological pictures;

(b) geometric space as a subdomain (locations and geometric individuals): many visualizations in the domain of GIS (Geographical Information systems) belong to this type (cf. also Fig. 89). The distribution of invisible attributes (e.g., vibration) on the flatly unfolded surface of a space shuttle during entering the atmosphere (for example) is of this type, too;

(c) a dense one- to four-dimensional subdomain apart from geometry: many state spaces in physics provide such subdomains, e.g., frequency spaces of Fourier transforms, or impetus space in Hamilton mechanics;

(d) color as a subdomain: this case is often, but not necessarily combined with one of the above: take for example the graph of a "color mountain" in isometric perspective where the height encodes the size of the area that could be printed in a certain color with the stock of pigment still available in the printer;

(e) none of the above. In this rest category, certain target concepts have to be seletcted to be matched to location as the pictorial base structure.

Apart from case (e), the subdomains mentioned above have a canonical projection into pictorial syntax that can be directly used as the core of the visualization. In particular the geometric spaces – either the projections of sortal objects as in (a), the genuine geometric domains in (b), or the quasi-geometric dimensions of (c) – map into the spatial substrate of pictorial syntax while other properties are more likely to be associated with elementary pictorial marker values (color), their grouping into pixemes (marks), and the high-level properties of associated objects (length).

On this general level, we can already see that the primary »picture content« for structural pictures has to be a bit more complicated. It contains in general a combination of the geometric concepts governing the pixemes as the perceptual basis, and the target domain. The corresponding concepts of both fields integrated in a particular instance of »picture content« may be directly linked, or they may be linked by means of sortal con-

cepts as mediators. That is, picture content for structural pictures contains an association of concepts from one, two or three fields. Let us have a short look at some of the possible combinations:

[1](a): »picture content« corresponds to representational images with an additional projection from some pictorial marker values to non-visible attributes of those sortals.

[2](a): this is essentially the case for representational pictures – if we want to understand them as a special sub-type of structural pictures: »picture content« contains sortal concepts with their constituting relation to geometric Gestalts.

[2](b): this covers the special case of purely geometric pictures – the semantic relation deviates to an identity relation, and »picture content« consists of entries from the geometric field only, without considering a constituting relation: the syntactic structure is meant literally. This type of visualization plays an important role when we consider reflective pictures: the semantic reduction can be used to focus exactly on that aspect in pictorial communication (cf. Sect. 4.4.5.1).

[2](c): similar to representational pictures, »picture content« contains in this case concepts with an association between two fields: in contrast to the representational pictures, the association is not an object constitution, though it has a similar effect: it provides the target concept metaphorically with a visually perceptible component.

[1](e): this is certainly the most complicated case, since »picture content« has to deal with a double projection: the concepts of sortal objects are contained together with their geometric projection (as the perceptible basis) and a metaphoric relation projecting the sortal structures onto a third field.

In order to ease the beholder, those metaphoric connections have to fulfill some minimal semantic criteria: a close structural correspondence between source domain elements and target domain elements.

4.4.4.2 Finding Appropriate Visualization Parameters: An Overview

Attributes of the target domain are usually divided in the following classes: they are qualitative or quantitative; and they are ordinal or nominal, i.e., with or without an inherent order that may be linear, cyclic or semi-ordered. Quantitative attributes are furthermore divided by their dimensionality. The names of the months, for example, are qualitative and ordinal of the cyclic type; the kind of trees found in Yosemite is (viewed as) qualitative and nominal (at least if we ignore the alphabetic order that could be imported by the names of the trees). Temperature is a one-dimensional quantitative property – like all quantitative properties, it has an inherent order, which is linear in this case. An analogous distinction applies to the attributes of the source domain, in particular of the geometric field providing pictorial syntax. Shape, for example is not quantitative and has no inherent order, while size is quantitative with a linear order. Color *per se* is a three-dimensional quantitative attribute without an ordering; however, hue is a one-dimensional quantitative feature with a cyclic order, while intensity is also one-dimensional but linearly ordered. Color categories (e.g., the naming of colors) is qualitative and often has a conventionally determined linear or cyclic order: from the

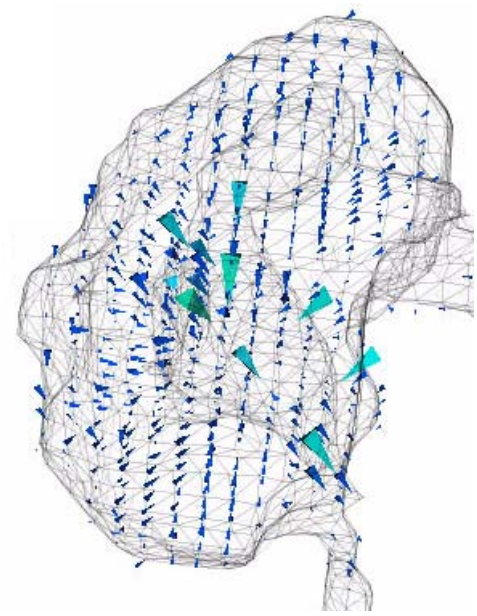


Figure 90: A 3D Grid of Arrows Indicating Flow Gradients

colors of the rainbow, through the color code used in cartography or the different one from anatomic sketches to the formal descriptions of the color of wine in oenology, the categories of human hair color or the dichotomy between cold and warm colors. Note that such color categorizations in the beholder's background may interfere with the intended interpretation of a color encoding.

Quite obviously, the primary choice for the elements of a metaphoric transfer is given by attributes of the corresponding class. For example, quantitative properties of a certain dimensionality and ordering should be expressed by pictorial attributes of the same kind. Accordingly, size or saturation should not be used for nominal target properties because they will probably be perceived as ordered, and thus induce an erroneous understanding. Varying shades of gray (or a monochromous scale) with their inherent linear order correspond better to linear quantities than color. IGNATIUS & SENAY [*1994*] provide a collection of heuristics.

So far, finding appropriate visualization parameters does not seem too complicated. But the main problem of visualization is connected with the amount of data to be visualized, and the high number of properties to be used simultaneously (multivariate visualizations). The source domain only offers a limited number of properties, and not all of them fit to the target properties under investigation. Even if an appropriate association could be found, the grouping in one dimension leads to interferences with the recognition of other dimensions in our perceptual apparatus. For example, hue dominates form, i.e., it may mask a form distinction; but it is itself dominated by intensity [HEALEY *2000*].

An elegant and often-used method for composing more target properties into a visualization is given by means of icons (or glyphs). Those are two- or three-dimensional entities that group several target properties by means of attributes such as shape, size, color, and position. They appear either as embedded images (icons) in the image plane or as 3D-objects in the picture space. The distribution of sociologically important features across the cities of a country can, for example, be encoded visually by "smiley faces" placed at the position of the cities in a map: the size of the head circle indicating the tax volume, the filling color from a spectrum between blue and red representing the average amount of days taken off by illness, the curve and size of the mouth line encoding the direction and amount of contentment of the inhabitants, etc [CHERNOFF *1973*].

Arrows and weathervanes are prominent members of the glyph family. They are examples of a special class of glyphs that are also used for gaining an overview about high amounts of data items: think of a streaming pattern graphically indicated by a dense field of arrows (Figure 90). Stick figure icons as shown in Figures 91 and 92 are a more complicated example. Here, the limb angles of an idealized stick figure serve the visualizer for encoding target parameters – color and thickness of the limbs could be employed for representing additional target properties, as well. If the data items are densely

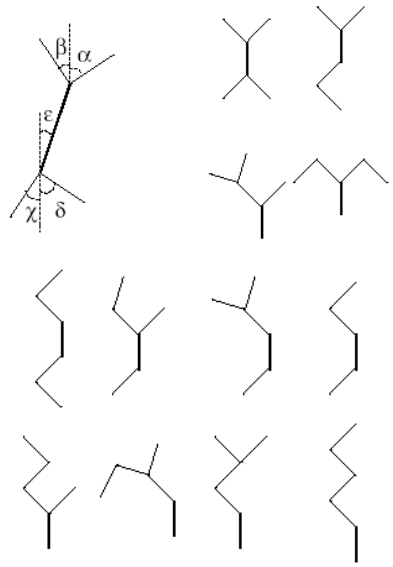


Figure 91: Stick Figure Icon with Five Attribute Ranges

Figure 92: "Data Texture" with Stick Figure Icons

packed on a two-dimensional grid indicating location or other target dimensions of according kind, the stick figures visually melt to complex textures in which differences can be easily perceived. The association of a difference to the parameters causing it is, however, not really "readable" from the graphic. Furthermore, it is often unclear whether different parameters may lead to the same texture.

If the amount of data items or target domain attributes becomes extremely large, even glyph textures and similar techniques that lead to a single static visualization meet their limitations. The visualization has to be divided among several pictures presented at once side by side, which is traditionally managed by means of the Model-View-Controller paradigm: the target domain is a model certain views of which are generated depending on control parameters determined by the user. LAU ET AL. [*2001*] present alternatively a metaphor for multi-screen visualizations derived from theatre that is particularly well suited if many users work with the data and its visualizations: the target entities with their properties are conceived of as actors that show different roles on several "stages" – a small amount of separate images – according to "scripts", i.e., visualization plans provided by the directors (users). A virtual stage manager coordinates the different scripts for all actors and each director.

4.4.4.3 Interactive Visualizations

But even the number of pictures to be presented simultaneously without confusing the beholder is quite small. Using time to divide several aspects of a complex visualization has proved to be more intelligible. Animations and – even better – interactive pictures provide a much more general means. Animations add just one dimension to the pictorial base structure that could be employed for encoding an orderd property of the target domain. With interactive manipulations, various target properties can be bound to vary with the animation's temporal visualization axis. With the slicing technique, for example, a user can change the values of one parameter (i.e., bind it temporally to time) and gets successive projections of the other aspects of the complete whole data set (Fig. 93).

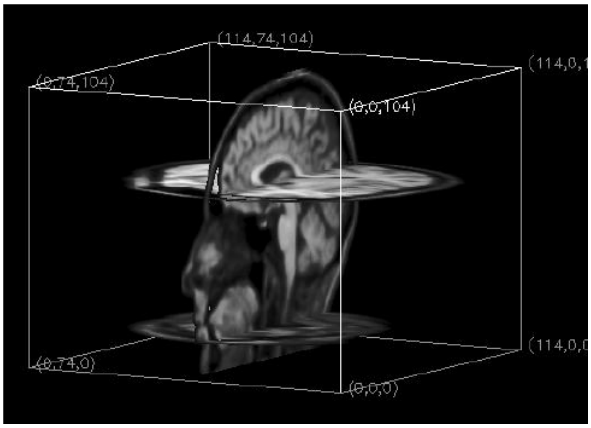


Figure 93: Interactively Slicing Views of the Data Set

In general, viewpoint control techniques are an important aspect of interactive visualizations in order to help the user keep an overview into which he or she can integrate the partial views currently focused: we have already mentioned fisheye zooming, i.e., the use of a bend perspective (Sect. 4.3.1.4). Generally, non-linear projections can be used to provide visualizations with a detailed area of focus (interactively changeable), a low resolution surrounding, and a more or less continuous transition area.

Most of the criteria mentioned so far are concerned with the correspondence between the metaphorical content and its graphical form – we have thus given a short recapitulation of the semantic aspects of structural pictures (that had been omitted in Section 4.3 on purpose). In principle, the effectiveness of the semantic criteria for selecting appropriate visualization parameters has to be controlled by a beholder model, as well.

We have reach here an interesting distinction in the use of an interactive system: In the beginning of this section we have mentioned that visualizations play an important role for finding good metaphors for the data to be visualized. The metaphoric mapping is used for helping to build a constitution of a field of concepts the internal structure of which is still highly unclear. To that explorative purpose, the visualizations are often employed in a strictly soliloquial manner. But visualizations are highly valuable tools for coordinating knowledge and action between individuals, too.

4.4.5 Remarks on the Pragmatics of Computer Art

> *Das Prädikat Kunst ist ein soziales. Der soziale Prozeß der Rezeption macht ein Werk zum Kunstwerk, nicht der Schaffensprozeß des Künstlers. .....*
> *Kontrolle über den Prozeß der Bildproduktion wird bewußt. ..... Das heißt aber nichts anderes als: bewußt werden Zeichen über die Produktion zum Gegenstand der Arbeit gemacht.*
>
> [NAKE *1996, Sections 8 & 7*]

The third general category of pictures distinguished by SACHS-HOMBACH is the one of reflective pictures. Pictures that are not used in the primary sense of showing their content but instead of demonstrating aspects of pictorial communication are called reflective, as has been explained in section 3.5.5. Thus, the characteristic distinction of that category is not one of semantics but of a pragmatic nature: it is a different reception mode in which the picture is dealt with. Considerations on the relations between reflective mode of use and reflective pictures therefore start this section.

Reflective pictures do not play a prominent role in the practice of computational visualistics on the first view although they are widely employed in textbooks on computer graphics, information visualization, computer vision or any other domain dealing with computerized images. They are used to exemplify the algorithms that have produced them. There is, however, one field of generating or manipulating pictures by computers that is directly concerned with images to be received in reflective mode: the field of

computer art – or rather of art with the computer as a tool. A short discussion of a few characteristic examples forming a coarse sketch about the categories of aesthetic considerations in the context of computer art completes the overview.

4.4.5.1 Reflective Pictures and the Reflective Mode of Reception

Let us start with a simple thesis: any picture can be used as a reflective picture. Indeed, we only have to "quote" that picture in order to communicate about the particular details of its process of creation and its history of reception. It has been mentioned before that screenshots from computer-generated or computer-manipulated pictures are frequently employed in papers and textbooks decribing some aspect of computer visualists' work. They are not meant literally: showing diverse variants of a certain teapot does not intent to move the beholder's focus of attention to teapots (or one particular teapot), but to various forms of, for example, shading calculations.

Representational pictures and structural pictures differ in the way the syntax is related to the semantics; reflective pictures differ from both of them by a different attitude of the beholders. This special attitude has been called 'the reflective mode of reception' in section 3.5.1. In this mode, we show ourselves a picture as an example of one or the other of the many aspects of pictorial communication. Indeed, this is what we usually do when visiting an art museum, and pictures of art can generally be interpreted as pictures that are made specially for being received in reflective mode. In the motto of this section, F. NAKE, one of the first studying and performing art in the context of computer science, insists on exactly this point and its two aspects: to conceive of reception as constituting the picture as a reflective one; and to take the creation of pictorial signs as the subject of reflection, which can only mean the constitution of the situation of communication in which the sign partakes, not just the material production of the image vehicle.

The context built by a reflective picture therefore contains not just what the »picture content« determines, but a communicative situation with that picture, its vehicle, and a generic sender and receiver. The particular instantiation of that situation is, then, considered, analyzed, put in relation with instantiations evoked by other reflectively used images. The narration (in the wide sense) provoked by such pictures obviously involves the discourse of art reviewers, art historians, and art scientists. They thus deal with the same subject as general visualistics but from the particular perspective gained by studying reflective pictures, not pictures in general. Art is not restricted to reflect only the primary modes of picture reception. Reflecting the reflective mode however increases the complications drastically and usually leads to contexts so complicated that only specialists seem still able to cope with. Some of the problems contemporary art has with reception among the non-expert public may root in the complexity of reflecting the reflective mode that is involved in their pragmatics.

Nevertheless, quoted pictures as well as artistic pictures have an underlying "direct" use – they are also representational or structural pictures that could be used in a non-reflective manner. The portrait of a certain merchant created by REMBRANDT can be employed in analogy to a passport – ignoring thereby all the sophisticated considerations of art history or art science that deal essentially with what the reflective mode of reception may reveal from that portrait.

Could there be pictures that can *only* be received in reflective mode – essentially reflective images? We can only speculate at this place: let us have a look at non-figurative art, and let us conceive of it as a special kind of visualization: the visualization of colored geometric entities where the structural isomorphism degenerates to the identity re-

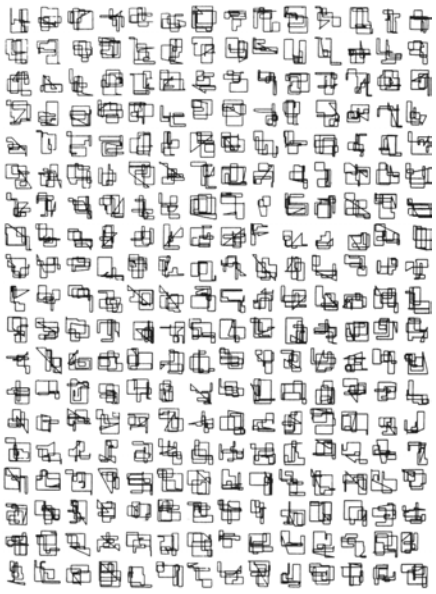


Figure 94: *23-ECKE*

GEORG NESS *1964*

Figure 95: *Geradenscharen Nr. 2*

FRIEDER NAKE *1965, 50 * 50 cm*

lation (i.e., case [2](b) in section 4.4.4.1). Such a picture regularly resists any figurative (sortal) interpretation. Due to the missing color legend, even a conventional interpretation as a geometric metaphor of some other subject fails. Thus, the image urges the beholders to "go into reflective mode": a failed attempt of communication regularly activates the process of reflecting the conditions of communication (basically in order to find the reason for the failure). In consequence, it is highly probable that a beholder of a purely geometric picture tries to find a reflective interpretation in the context of pictorial communication – following, for example, the line of argumentation that no pictorial sign act can do without geometric Gestalts, or that certain geometric configurations evoke spontaneous emotional reactions, etc. That is, structural pictures of the case [2](b) without a legend can only be interpreted in reflective mode: they are to be conceived of as essentially reflective pictures.

No wonder then, that the first works of what is called 'computer art' is of this category, and not figurative (representational). It is often mentioned that the technical restrictions of that time are responsible for the non-figurative character of the works exhibited by NAKE, NEES or NOLL 1965 in the first ever arts exhibitions of pictures created with the computer (Fig.'s 94 & 95). The considerations above add, however, the argument that pictures of that abstract kind have an inherent advantage of being perceived as art, and not as mere by-products of testing new technical devices.

Quite obviously, the attempt to create works of art by means of a computer and its peripheral machines has to be markedly distinguished from the use of systems for (re)producing pictures in a certain style or varying the representation style of a given picture. Algorithms for such tasks are regularly included in commercial paint programs, and still form a "hot topic" for developers of non-photorealistic computer graphics.

4.4.5.2 Computer Art – Art with the Computer

There is a general agreement in that the term 'computer art' is not well-chosen [STELLER *1992, 11*]: the computer is essentially the artist's instrument but does not determine a specific style *per se*. With the words of one of the pioneers, H. W. FRANKE

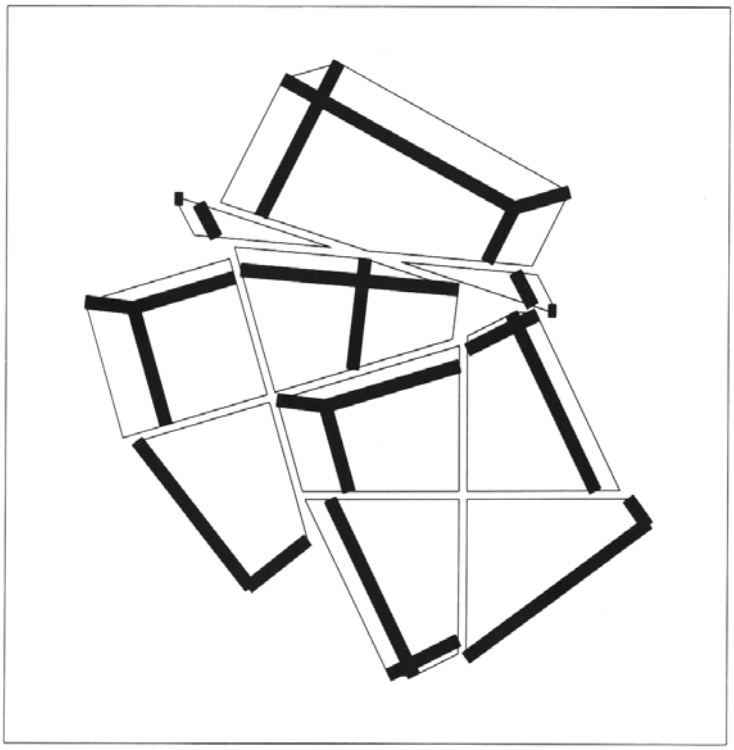


Figure 96: P-361-E

MANFRED MOHR *1984* (acryl on canvas, 120 * 120 cm)

[*1987, 335*]: *The term 'computer art' refers neither to a specific style nor to a particular quality, it merely characterizes the instrumentarium.*

One of the most consequent artists using a computer as his tool is the German MANFRED MOHR. Working essentially on a single theme, the cube, he takes up, we might say, the pictorial discussion of the cubists in the early 20th century about the relations between the multi-perspective sortals and their traditionally mono-perspective depiction. Like the works of BRAQUE or PICASSO, MOHR's pictures are in fact figurative. They are derived as fragments of cubes, each fragment viewed from a different perspective (e.g., Fig. 96).[79] Usually grouped to whole sequences of pictures, their very reduced syntax essentially is determined by means of the computer and then manually painted on canvas.[80]

Those pictures can be read as an offer to discuss on a very abstract level the role perspective, or more precisely: a set of multiple perspectives, plays in our communicative behavior [KING *2002*]. It may also be employed to focus a discussion about the role of algorithmic processes for creating images, a line followed, for example, by NAKE [*1996, Sect. 9*]:

> *The signs of MOHR's pictures (the name of which are like P197-H or P370-P) stand for themselves, but they also represent something different: the four-dimensional situation*

[79] 'Cubes' here has to be understood in a more general sense, as MOHR started with ordinary three-dimensional cubes, but has proceeded to their four- and even five-dimensional counterparts, the hypercubes – that is, geometric entities that, like sortals, integrate conceptually several (3D and 2D) projections.

[80] That the pictures are part of a series is in fact an important hint to understand the common figure all the pictures of the series refer to, in particular as fragmentation and the multi-perspective views often obscure a direct recognition, and as hypercubes are objects not too familiar for recogition.

Figure 97: Picture Generated with the Fully Automatic Online-Version of COHEN's *Aaron*

> *indicated. That different thing is mostly unknown, unviewed, invisible. By means of the program, MOHR makes visible some aspects of that mathematical reality. Contingency and arbitrariness of each single one of his signs is bound together by the algorithmic uniqueness.*

The two lines of argumentation do not contradict or exclude each other, since each is a legitimate reflective use of those pictures in a corresponding communication of art lovers. The artist, of course, may prefer one or the other himself. For MOHR, who has founded together with others the artists group *Algorists* few years ago, "the shift from uncontrollable metaphysics" (of beliefs about objects and perspective, we may interpret) "to a systematic and logical constructivism" (a reasonable understanding of the relation between geometry and the sortal field, that is) "may well be a sign of tomorrow" [MOHR *1976, 96*].

Quite a different result of integrating the computer into the artistic considerations is reached by British artist HAROLD COHEN (Fig. 97). Since about 30 years, the traditionally trained and experienced painter works in developing and programming a system called *Aaron* that would be able to generate autonomously pictures in a way that simulates up to some degree human cognitive processes relevant for drawing (and later painting). *Aaron* is constructed as an expert system containing as its knowledge base rules about generating images on syntactic and semantic levels.[81] There are rules for determining unused space, for drawing closed or open figures, or for controlling repetitions. Other rules select a pose for declaratively given stick figure models, expand the models to provide them, so to speak, with flesh, or differentiate plant models to a par-

[81] COHEN prefers to see *Aaron* as an "expert's system" rather than an expert system: the latter provides non-experts with a "petrified" version of an expert's argumentation; the former helps an expert to understand what s/he is doing; cf. [COHEN *1988, Appendix: "Conclusion as delivered"*]

ticular instantiation of branching, branch thicknesses and leaf forms. On each level, variances are determined randomly.

F. NAKE summarizes COHEN's approach as follows [*1996, Sect. 3*]:

> *COHEN, the painter, urges the machine to do what he wants. He does that by exteriorizing parts of his drawing and painting. More precisely: he externalizes a part of his knowledge about drawing and painting. A part of his thinking, that is; the part that he can put in algorithmic form. HAROLD COHEN has gone further than anybody else in this world to fix a certain kind of form design and paint application in a rule base. Those rules are so precise (unique) and contain as much openness (arbitrary) that the computer can follow them, and appears to become creative.*

So what is the discourse initiated and focussed by means of COHEN's "*Aaron* pictures" if we take them as reflective images? What context do they build? Recall at this point again our interpretation of data structures and algorithms (sequences of operations) as formalized tools in rational argumentations of sections 2.1 and 4.3.1.2. *Aaron*, then, contains a formalized version of arguments that can be used when discussing about image making and its cognitive prerequisites. The pictures produced by the system can therefore be read in the reflective way as exemplifying corresponding parts of those argumentations. There is, then, a closer relationship to quoted images in a textbook on computer graphics than is the case for MOHR's pictures. *Aaron*'s products provoke the reflective narrative if we know that COHEN has generated them in that specific way, while MOHR's fragmented hypercube projections can be discussed along the same lines without mentioning the tool utilized.

Of course: COHEN states explicitly, that *Aaron* is not intended as a simulated artist producing works of art. He understands the program as an assistant, a highly developed and partially autonomous tool, that is – but still a tool dependent from the context of use. Since the artist also understands art as a communicative act that has essentially a pragmatic embedding unavailable for the computer, the pictures generated (proposed?) by *Aaron* are still COHEN's pictures, because he is the one binding them originally into a sign act.

4.4.5.3 Interactivity in Computer Art

> *A statement on computer-assisted visual arts that is often meant as a criticism is true: basically every static image presented on the monitor could also have been created in a conventional manner, the only difference being the time needed to produce it. Therefore, the significance of computer graphics lies only in animation.*

This quote from FRANKE's programmatic essay of 1987 on the (then) future of computer art leads us not only to animations, but especially back to interactive pictures, which we have already identified earlier in this chapter as the crucial contribution to the subject of visualistics from computer science. The serial productions of images, which always has played an important role for artists working with the computer (including MOHR and COHEN), is condensed into a continuous animation that can take into account parameters provided by the beholders. Here, the question arises whether such forms of pictures have to be understood in the manner of the media of class *IV*, or rather in a different way. Let us look, for example, at the specific form of interactive pictures used in web art, where many beholders can simultaneously interact with the same picture-generating schema. While tele-rendering in media of class *IV* in the strict sense provides a bundle of potential pictorial sign acts from which the beholder selects – directly or in-

directly – one or the other, web art essentially consists of one unique though complex sign act that explicitly integrates the beholder.[82] Two users of an interactive anatomy textbook can legitimately state that they had encountered different sign acts, whereas two receivers of a project of web art, despite having different experiences, have participated in the same sign act (or we would abandon the notion of a unique work of art altogether).

Indeed, this type of images is a special derivation of the generic concept »image«, with which the western tradition of art history is not too familiar. For a long period, the identity of an artistic image has been conceived as bound to one particular picture vehicle, the *original*, which was thought – for example due to the density of pictorial syntax – to be not fit at all for proper copying. Recall however at this place the remarks in Section 3.1 about religious sand or bark pictures in American and Australian indigenous cultures: their vehicles are destroyed immediately after use, yet the pictures are said to be the same in different actualizations of the ceremony. The identity criterion of this concept of image has a resemblance to the one of music: a generative core – externalized as a score – can be instantiated again and again as the same music on different vehicles even with slight alterations. Recall, too, the generative picture format mentioned earlier: in section 4.2.1.1, the example of fractal pictures with zooming operation was given – a unique computational schema directing how it can be instantiated by different beholders. The generic concept of pictures encompasses a sub concept where the picture is fixedly bound to a certain individual picture vehicle, as well as a sub concept with an elaborate two-level conception of identity similar to the "score-performance" distinction in music. During the 20th century, the later sub concept has become more dominant in Western art, in particular with the employment of corresponding technical tools like video or the computer by the artists.

Reflective uses are more or less obvious for generative web images, as well. HUBER [*1997, 188*], for example, mentions in a survey on web art a piece of JOHN SIMON JR. that certainly evokes in the beholder (or the reader imagining it, in our case) the discussion on syntactic properties of pictures in section 4.2.1:

> *In a second work for the web from JOHN SIMON JR. titled* 'Every Icon' *(http://stadiumweb.com/EveryIcon), a Java applet generates all combinations of black and white squares. The work runs since March 1, 1997. It can be viewed only in its beginning – a computer has to run with that little application day and night for years.*

In fact, about several hundred trillion years are necessary, SIMON points out on the commenting web page, for the program to generate systematically all variations of the 32 * 32 pixel matrix used on the way from completely white to completely black.

More complicated versions of interactive computer art are derived from highly immersive systems. A prominent example is an installation of the Canadian CHAR(LOTTE) DAVIES, first exhibited in 1995:

> Osmose *is an immersive interactive environment, involving head mounted display, 3-D computer graphics, and interactive sound, which can be explored syn-aesthetically. On a second level, the installation offers visitors the opportunity to follow the individual interactor's journey of images through this simulacrum of nature. With the aid of polarized glasses, they watch his or her constantly changing perspectives of the three-dimensional image worlds on a large-scale projection screen. The images are gener-*

[82] The integration of the beholder into the work has become a standard theme of art with modern media, see, for example, in video art the pieces of B. VIOLA.

Figure 98: Watching the Beholder: the *Osmose* Installation

> *ated exclusively by the interactor, whose moving silhouette can be discerned dimly on a pane of frosted glass.*

This description of GRAU [*2003, 193*] already indicates that DAVIS's installation has indeed two types of beholders: the single interacting "immersant", and a relatively passive audience watching simultaneously the immersant's slightly grotesque shadow behind a screen – with the cables, the head-mounted display, and a sensor vest registering the immersant's breathing and other movements of the torso as important parameters of interaction – and a video projection with exactly the computer animation produced for the immersant in real-time depending on his or her movements (Fig. 98). For the immersants, the concoction of reactive *trompe l'œil* and animated sculpture provided by the computer and its immersive interface inhibits the symbolic mode to a very high degree. They regularly do not perceive pictures anymore, but a strange and interesting situational context: that is, they experience the pure deceptive mode.[83] GRAU explains [*2003, 195*]:

> *Like a scuba diver, the observer floats upward with lungs filled with air, whereas regular breathing produces a calm and balanced state. Divers are well acquainted with the feeling of immersion, the physical experience of being completely enveloped and slowly floating through the watery element. Not surprisingly, it was being underwater that gave CHAR DAVIES (who is a passionate diver) the inspiration for this finely gauged, physically intimate synthesis of the technical and the organic. Because the interface technique of* Osmose *utilizes intuitive physical processes, the observer's unconscious connects to the virtual space in a much more intense way than with a joystick or a mouse.*

[83] There are, however, integrated a few fractures of the illusion – in GRAU's description [*2003, 194*]: *Two textual worlds serve as parentheses around this simulacrum of nature: The 20,000 lines of program code for the work are visible in the virtual environment, arranged in colossal columns; and a space filled with fragments of text-concepts of nature, technology, and bodies, all penned by thinkers, such as BACHELARD, HEIDEGGER, and RILKE, whose ideas were untouched by recent revolutionary developments concerning the image.*

Figure 99: *Osmose* immersant with head-mounted stereo display and sensor vest

Figure 100: Still from *Osmose* ("Subterranean Rocks and Roots")

The audience – being presented something like a blurred version of Figure 99 together with the stereo-version of Figure 100 – can more easily gain the distance of the immersive mode of reception necessary for perceiving something as a picture. Not allowing the audience to see the immersant directly but only his silhouette projected onto a screen – put in frame so to speak – is another strong indicator for perceiving the installation as a coordinate set of two animated pictures. Observing the explicit opposition of the viewer and the viewed, of the interactor and the resulting image, in fact suggests for the audience to use the reflective mode. The deceptive mode of the immersant and their own immersive attitude towards the computer-generated picture opens a discrepancy for the beholders that prepares them for discussing the dual nature of pictures, the spontaneous deceptive mode (the picture's perceptoid character), and the distanced symbolic mode (the picture's sign character).

Most interestingly, the published critics deal often with the deceptive mode of the immersant only. Not questioning the extraordinary experiences gained as immersant, many reviewers ask – quite with good reason – whether this really is art already, or simply kitsch. Most comments (including remarks of DAVIES), change between technical details and more or less esoteric evocations of an immersant's state of mind:[84] they are not too helpful to decide the problem, which we shall not elaborate here. However, the question obviously does not arise for the installation viewed as a whole from the audience's point of view.

HUBER [*1997, 187*] distinguishes reactive, interactive, and participative projects of computer art. *Every Icon* certainly belongs to the (in that dimension) most primitive kind of reactive work, since the beholder can do only one thing: changing the internal clock of the computer, far from the real potential even of reactive pieces. *Osmose* is clearly representative for fully interactive projects. Participative projects in the full

[84] cf., e.g., [DAVIS *1998, 56ff*]: *Osmose is a powerful example of how technological environments can simulate something like the old animist immersion in the World Soul, organic dreamings that depend, in power and effect, upon the ethereal fire... Osmose also reminds us how intimate we are with electronics, in sight and sound, in body and psyche.*

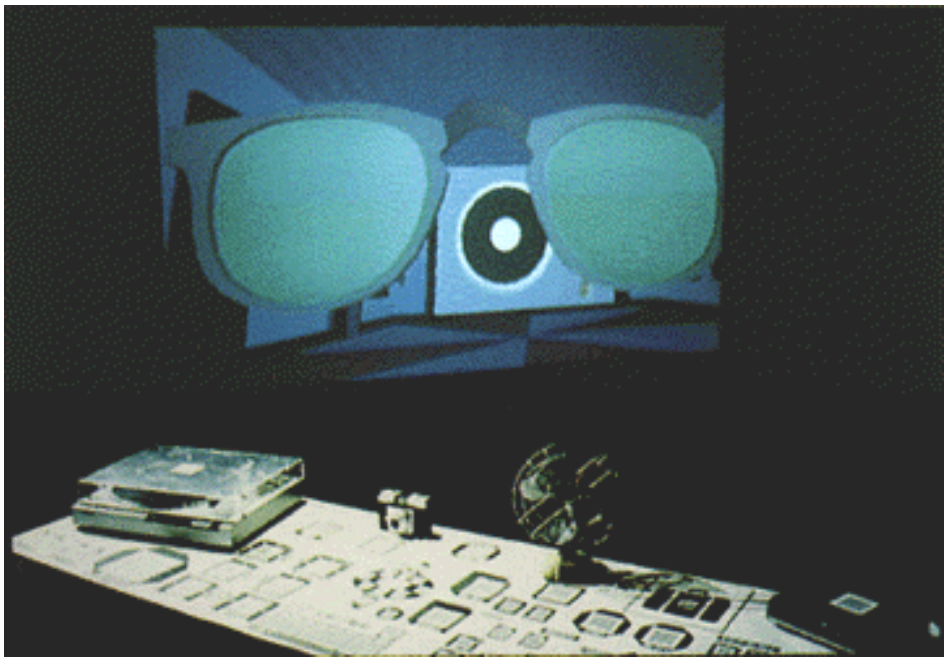


Figure 101: Photography of HOBERMAN's *Bar Code Hotel* Installation: objects with bar code on the contol panel with bar code reader in the foreground, computer-generated picture in the back

sense exceed interaction since they additionally allow the users to modify the generative schema (or the bundle of messages) – mostly by actively increasing them.

PERRY HOBERMAN uses in the participative creation of *Bar Code Hotel* a symbol typical for our economy: the bar codes found on consumer goods. The participants have to scan the bar codes of given objects in order to select computer-generated object models to be rendered and projected stereoscopically on a large screen (Fig. 101). The picture shown is thus a result of the cooperative effort of all the beholders.

The basic principle of participative artwork is closely related to the art form of "happenings" popular in the 1960's and '70's mainly by members of the Fluxus movement starting with GEORGE MACIUNAS in 1962. JOSEPH BEUYS was one of its prominent artists. Fluxus is still another form of visual art incorporating its beholders to a much higher degree than traditional pictures do. In a typical happening, however, the beholders/participants are present in the flesh. They cannot avoid showing signs of their spontaneous reactions (e.g., while being drained with the blood of a freshly slaughtered animal – recall the œuvre of Austrian artist HERMANN NITSCH: the "Orgien Mysterien Theater"). And this is precisely what the artists want to focus on.

Multi-user art operates in contrast with immaterial representatives, and hence with a marked distance between the participant and the events happening to this representative. The corporeal (non-)existence of the users and its necessity as a prerequisite to communication is often a major theme for multi-user artists, too. Since the aspect of technical mediation seems on first view to disappear in communication in immersive multi-user environments – a medium of class *III* allows the users seemingly to communicate as in a medium of class *I* – reflections of this deception are often evoked by immersive participative art projects.

A weak form of participation is realized if the immersive context can be shared simultaneously by many users, though not modified in its generative schema. As an example for such a work of art, the project *Technosphere* of a team around the British media art-

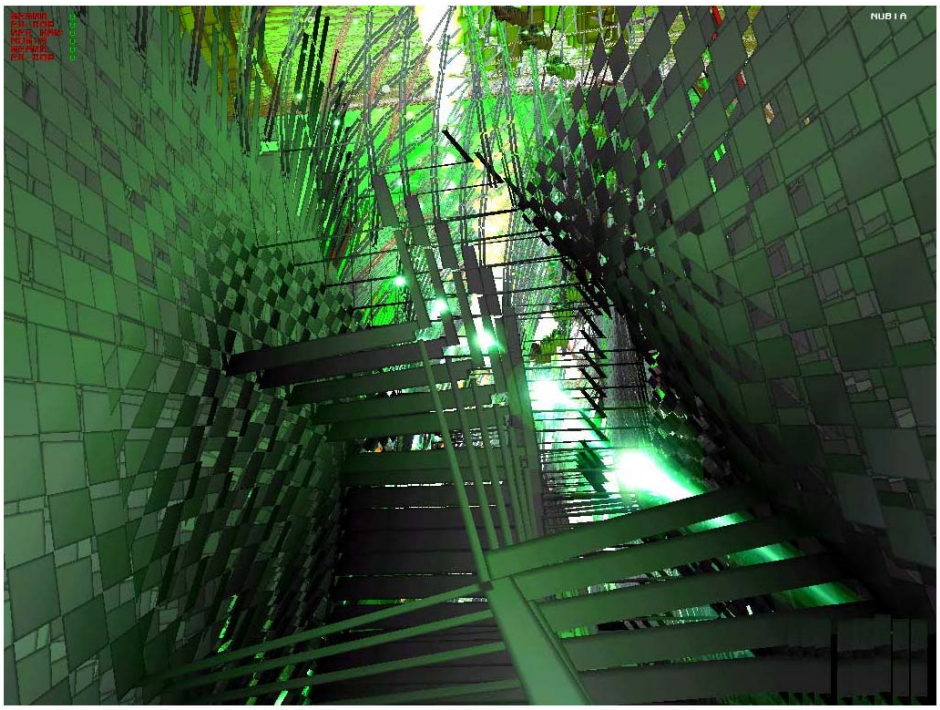


Figure 102: Screenshot from *Peacekeeper* by TOM BANKS (alias NULLPOINTER)

ist JANE PROPHET is usually mentioned [REICHLE *2002*]: developed in several steps and launched online in the WorldWideWeb since 1995, it is exhibited permanently since 1999 at the National Museum of Photography, Film & Television at Bradford, UK, in a version adapted for that presentation situation.

In the mathematical realms of a fractal landscape meant to stretch over 16km^2 dwell creatures whose appearance (and thus indirectly, whose behavior) is selected by the users of the system out of a given set of options. They are defined to move around, feed, hunt each other, reproduce, and die eventually. Significant events, in particular reproduction or death, are reported to the creating user by means of an email. On demand, a photo or an animation of the creature is rendered and sent to the user. In the web version, the creators of other creatures, with which one's own creature has interacted, can be contacted by email, as well. In the faster running museum version, real-time observation is possible from several terminals and replaces the email messages from the system.

Initially provided with 30,000 randomly designed creatures by the creators of *Technosphere*, several hundred thousand users have added over a million creatures so far. While it is not quite clear how the images generated in the Web version are employed to induce the reflective mode instead of interpreting them simply as representational images of fictitious contexts, the museum version offers again the option for the general audience to observe users interacting with the installation – though not as explicit as in the arrangement of *Osmose*.

The participative integration of the users within the simulated context is relatively weak in *Technosphere*, as users are present only by means of creatures that act rather independently from their creators. Avatars as more direct representatives are used, for example, in computer games, though the reflective potential of commercial computer games is quite minimal. Some artists have, however, begun recently to employ the technical options of computer game engines for artistic purposes. The American media artist LONNIE FLICKINGER, for example offers with *Pencil Whipped* a parody on the notori-

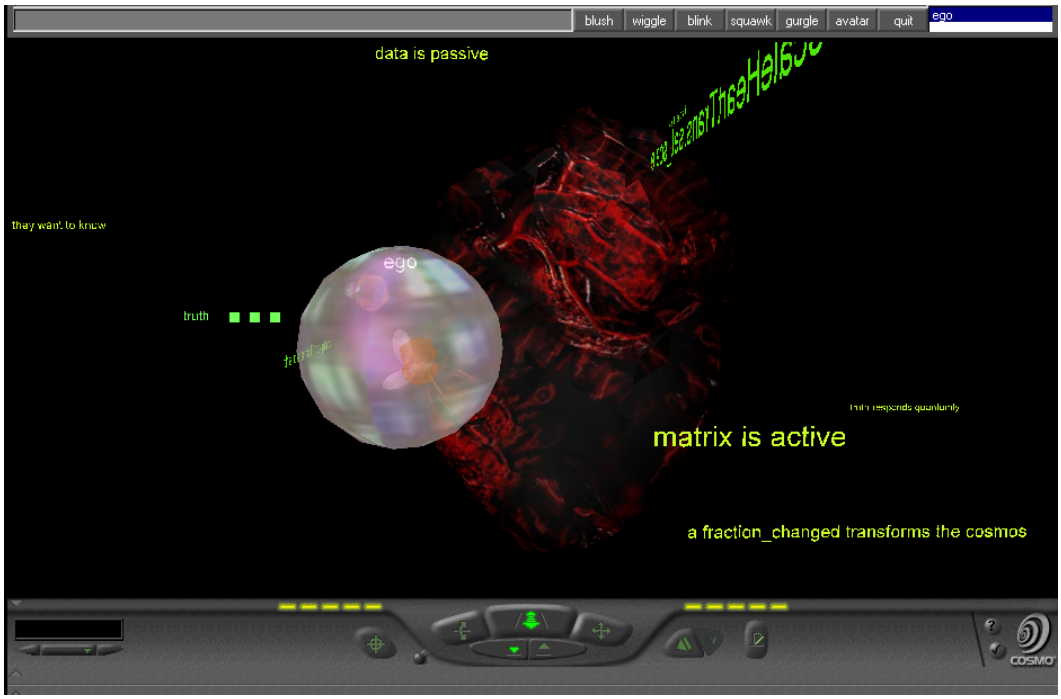


Figure 103: RACKHAM's *Empyrean*: View of the Context *Truth* with a pumping heart, moving text elements, and an Avatar ("ego") in the foreground; control panels for moving (bottom) and avatar animations (top)

ously naturalistic representation of brutality in ego shooters by replacing textures with a kind of coarsely sketched black-and-white comic drawings. Computer-controlled enemies are not modeled as 3D objects but as weird-formed flat papier-mâché creatures. The Dutch group JODI strips commercial action games like *Wolfenstein* [id Software, 1992] and *Doom* [id Software, 1993] from their naturalistic skin reaching strangely abstract labyrinths, in which the original game concept itself becomes quite obsolete.

The British artist TOM BANKS (alias NULLPOINTER) has presented a serious of game modifications based on the *Quake* game engine (Fig. 102). In his installation *Peace-Keeper*, two views of the modified game from the perspective of two adversary bots – computer-controlled characters – fighting with each other are projected onto opposing walls. The viewers – presumably hardly able to recognize corresponding actions in the disturbed images – can (try to) participate in the game and thus modify the flow of pictures while being watched by other visitors of the installation.

In these modified games as art, the user's presence as an avatar does not play too strong a role. Computer-mediated interactions between several users are also of no relevance. Australian digital artists MELINDA RACKHAM has gained attention with her multi-user project *Empyrean* (http://www.subtle.net/empyrean/): modeled as a VRML environment accessible by internet, users can explore several gravitation-free contexts: strange changing objects inhabit the unstructured virtual space, often half transparent, pulsing and with unclear boundaries, some with organoid forms – a pumping heart, a rotating eyeball, flickering neurons (Fig. 103). They float around each other emitting sounds; written words move in-between. Some forms act as portkeys to another context of the set. Since the text objects turn with the user's movements the fixed spatial orientations we usually expect are not valid: *There is no horizon to orient oneself to ... one must feel one's way around these immersing zones* [RACKHAM *2000*].

In contrast to *Osmose*, *Empyrean* does not work with the apparatus for a highly immersive experience like head-mounted stereo displays; moreover, navigation in VRML-contexts is not too natural and needs some practicing. Nor does RACKHAM count on beholders of the second order. Users can perceive each other visiting the contexts of *Empyrean* called Truth, Beauty, Chaos, Order, Charm, Strange, and Void. They are part of the picture, but merely in their representation by organoid avatars that mix with the other forms and are only distinguishable by the user's nicknames marked on them. Several buttons allow the user to activate certain animations of the own avatar (e.g., "blush" – a color change, "blink" – shrink and regrow) together with some sound effects that could indicate emotional gestures. Each user can also change the system's parameters so that he or she is able to watch the own avatar in a third-person view, i.e., to establish a symbolically mediated distance to his or her assumed body as the one being deceived by the artificial contexts of the system.

Empyrean can be seen as a work to be used for reflecting the difference between the pure deceptive mode of an immersive *trompe l'œil* and the symbolically distanced, intentionally controlled deception of picture reception proper. It puts a particular focus of attention to the role the bodies of the receivers play for those modes. In an accompanying text, RACKHAM explicitly relates her project to the immediate embodiment in a situational context prior to the constitution of a self that is aware of sortal objects, and the projection of that embodiment in the multiverse of symbolically mediated contexts [*2000*]:

> *As an avatar I simultaneously exist – internally and externally to the moist body, congruently within and with out, being self and other, both alone and networked. I am at once operating at binary oppositional points, co-present at both zero and one. Here I glide through spaces, I have no gravity, and my collision is false. This space reconstitutes my embodied self as soft object, with a resulting loss of rigid structural and symbolic self-definition, an amorphous material embodiment, a shifting node in a network constantly reinventing itself.*
>
> *And I am simultaneously hard edged. Physically located, constructed of zeros and ones, nothingness and singularities written along the x/y/z/axis of a 'real' located space. Hard space is where my collision is always true, where I bend slightly and my container often leaks, expelling viscous others that it has ingested.*

In a way, we have crossed with *Empyrean* a border of reflecting pictorial communication since pictures are here only part of a much more complex sign act including sound, text, the movements of the beholders and even the coordination of their interactions by means of a chat channel. Continuing that path of thought, we might find that cyberspace could very well be conceived of as a kind of super-perceptoid sign, integrating more sense modalities than the perceptoid signs known so far. The reflection arisen by visiting *Empyrean* is, then, much more a reflection on the use of that kind of sign than one of picture uses. Such a consideration does, however, exceed the frame of the current investigation.

* * *

With these fragmentary considerations on reflective images and computer art, we have finally reached the end of investigating pragmatic aspects of computational pictures, and thus also the end of our discussion of the generic data type »image«.

5 Case Studies: Using the Data Type »Image«

Having completed the trifold discussion on general aspects of the generic data structure that includes the type »image«, several applications now give an impression of how to apply the distinctions introduced. Individual tasks usually do not need all the aspects of the generic data structure; but the parts that have to be considered must often be transformed into more specific forms valid only for those pictures of the task at hand.

We start by describing a project on content-based image archiving and retrieval where the transformation from »image« to »picture content« is central (5.1). The second case study examines the inverse direction: an aspect of how to control a particular pragmatic aspect of the generation of rhetorically enriched pictures (5.2). Two varieties of using pictures in highly immersive systems are discussed as the third case study: the suppression of the symbolic mode in favor of a dominant deceptive mode of reception plays an ambiguous role in virtual architecture and for virtual institutes (5.3). Cognitive psychologists speaking about mental imagination indicate another kind of borderline case of pictorial signs. Computationally, the rules of speaking about mental images can be employed for modeling pragmatic effects in reference semantics by means of instances of »image« (5.4).

5.1 Semantic Requests to Image Databases in IRIS

A major problem arising in the present "age of the images" is the pure amount of pictures to be dealt with. The most elementary task of finding a certain picture or even a reasonably small group of relevant images in the enormous corpus of pictures available (for example, in a press archive) becomes increasingly hard. It is sometimes easier to produce a completely new picture instead – which then again contributes to clogging up the archive.

The classical tool for pictorial archiving is to index pictures by means of more or less arbitrary annotations associated conventionally with them. The keepers of the archives have to manually associate the annotations, essentially following their understanding of the pictures' essential features or contents and the principles of cataloguing of their profession. It is quite uncomfortable for any human being to describe the content of thousands of images following rather fixed criteria, and to construct a corresponding index. However, as soon as a new criterion becomes relevant, all images already categorized would have to be revised again: those processes are rather being done automatically.

In principle, we have to distinguish between several cases of image retrieval:

1. We want images that contain certain *syntactic features*, e.g., a red circle or a large patch of grass texture: Although such a request can be quite helpful if no other means of searching is provided, it is relatively uninteresting in most situations. If the archive is managed computationally sample pictures can be used to mark the features instead of giving them symbolically: IBM's system QUBIC provides exactly such queries by examples.
2. We want images that have a certain *picture content*, e.g., two persons in front of a forest: this is the most interesting case and we deal with it below. Specifying (partially) a picture content can reach from a single sortal object type ('a chair') to a fairly precise set of relative locations of several objects of certain types with associated visual features ('a red sports car with a blond guy sitting inside and a black-haired woman standing at the left side door').

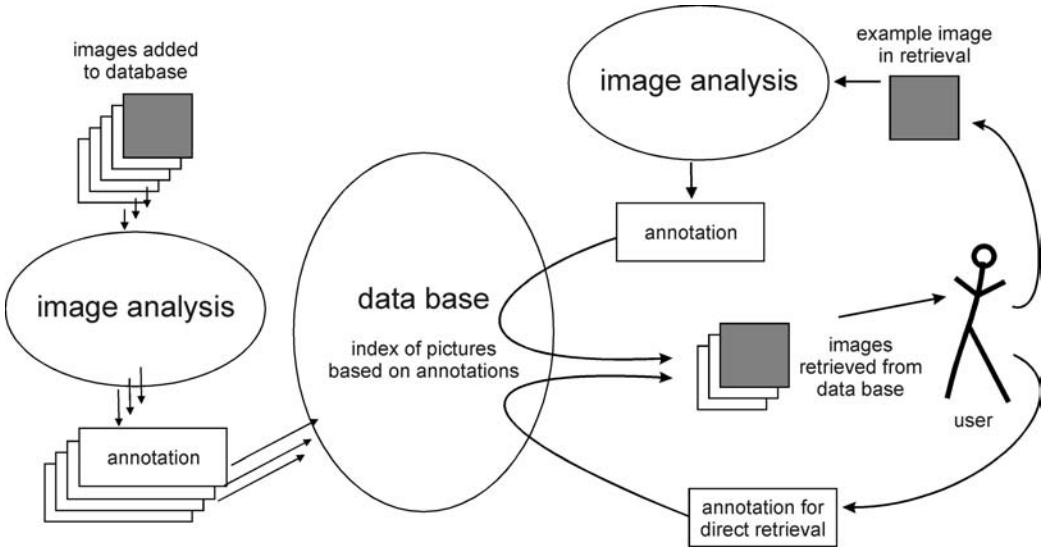


Figure 104: General Architecture of Content-Based Image Retrieval

3. We want images that have certain *individual referents* in them, e.g., a picture of the Taj Mahal. As was explained in section 4.3.2.4, image reference is problematic if not an unspecific individual is meant: 'unspecific' means that the individual in question is not known so far from some other context – it is just a spontaneously generated intentional object. In contrast to that, the specific individual pictured must be known as the same individual in other contexts, as well. The picture *per se* cannot establish such identification – another object with high visual similarity could be the referent just as well.

The task of retrieving pictures from a database in a semantic fashion, i.e., by means of giving content descriptions can be stated in relatively simple terms – nevertheless it is quite demanding to solve. Some aspects of such a task have been sketched already in section 4.4.1.4; but there, PINEDA assumed that descriptions of the images' content had already been derived in advance, and essentially "by hand".

5.1.1 Image Retrieval for Information Systems

The project IRIS (Image Retrieval for Information Systems[85], 1994–1996, Univ. Bremen) has approached the retrieving of a group of images currently of interest from a huge image archive, in which the pictures are automatically indexed according to their content (up to a certain level of detail). The system developed describes autonomously images by their content in a textual form. Only a specification of the general picture type is necessary, e.g., landscape picture, technical drawing or sports photograph, since the algorithms performing the image analysis depend on domain-specific parameters that cannot yet be extracted by the computer on its own.

The resulting annotations are fed into a standard textual database together with the reference to the corresponding picture files. A user of the system is able to retrieve the references to images by keywords from the annotations employing the well-known methods of text retrieval. The keywords can be derived by means of an analysis of a sample picture, as well (Fig. 104).

[85] The system has been implemented in C on IBM RS/6000 with AIX. It later became part of IBM's system "ImageMiner".

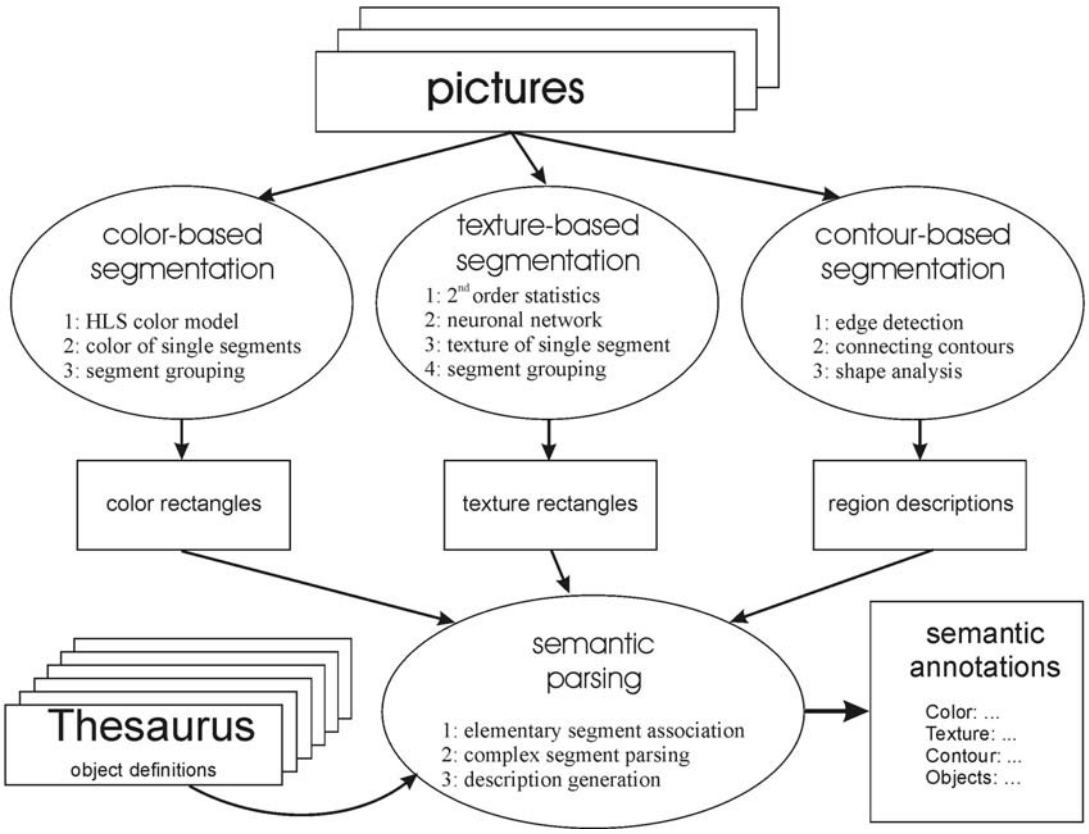


Figure 105: Architecture of Image Analysis in the System IRIS

The most interesting task is to construct the image content that is the basis of the annotations describing the pictures. An overview of this image analysis component is given in Figure 105. As described in section 4.3.2, the first step of picture analysis is to determine elementary pixemes: several types of marker values like colors, texture attributes, and contour elements are extracted from the image. Algorithms based on those described in [HARALICK *ET AL. 1973*] and [KORN *1988*] have been used for categorizing those features.

Elementary pixemes depending on color and texture are based on a grid with an adjustable size subdividing the image into grid elements. For every grid element, a color histogram is computed and reduced to a color category: the color category appearing most frequently defines the color of the grid element. Neighboring grid elements with the same color are grouped, and the circumscribing rectangles are determined. The results of color-based segmentation are described qualitatively by means of attributes such as relative size, position respective to the underlying grid size, and the color category.

Similarly, for every grid element, the system performs some matrix calculations getting some local statistic parameters like entropy, variance, correlation, and angular second momentum that base texture analysis. The mapping between the statistical values and the texture category to be used is performed by means of a neural net and depends on the type of scenes considered: certain statistic parameters may indicate one texture category in landscapes, for example, and another one in indoor scenes. Therefore, the neural net has to be trained in advance by backtracking with textures typical for the domain chosen (e.g., sky, clouds, sand, forest, grass, stone, snow, ice for landscapes).

Table 3: Original BNF for color rectangles in IRIS

```
<color description> := "HOR=<hor>,VER=<ver>,SIZ=<siz>,DIR=<dir>,
                        COM=<com>,COL=<col>"
  <hor>  := ll | left | middle | right | rr    ;; horizontal position
  <ver>  := uu | up | middle | down | dd       ;; vertical position
  <siz>  := XS | S | M | L | XL                ;; qualitative size
  <dir>  := Ver| Hor | Dec | Inc | none        ;; qualitative direction
  <com>  := Quad | Rect | Path                 ;; compactness
  <col>  := white | black | gray | red | yellow | blue | green |
            orange | violet | brown            ;; the actual marker values
```

Again, neighboring grid elements with the same texture type are grouped together so that the circumscribing rectangle can be used as the basis of the qualitative description.

Shape attributes are represented through contour-based region descriptions. Detection of edge elements based on the intensity gradient is a standard tool of image processing. To avoid the inherent scale-space problem of the gradient-threshold calculation, a "pyramid-structured" approach with several levels of resolution is used in IRIS. Relevant edge points (i.e., no noise) are collected into contours if they continue a contour hypothesis starting with the most prominent edge points. Closed contours are finally used to determine regions.

Color rectangles, texture rectangles and shape descriptions are encoded in a qualitative manner. Take as an example the BNF specification developed by the author for color rectangles given in Table 3.[86]

Spatial objects as the basic elements of the goal descriptions are associated to sets of segments that are visually perceptible in the picture, but they also involve relations to their parts, or to wholes of which they are parts. The definition of those part-whole relations for a particular type of object in fact organizes which sets of pictorial segments show an instance of that type, and which deviations are not to be rated as such instances. Spatial objects in the intended sense are constituted by the coordination of corresponding segments by means of object schemata relevant for the domain in question (Fig. 106). Context-sensitive techniques are used to guide this process. The goal is to eliminate ambiguity as early as possible by means of expectations.

An association between segments with the same marker values is usually only possible to elementary parts of spatial objects. To that purpose, topological relations between the pixemes found in the previous step are employed in a graph grammar parser to identify candidates for elementary parts of the scene in question [KLAUCK *1994*]. Thus, if a certain contour-based region, a white color rectangle, and a snow texture rectangle overlap widely, a region of snow is likely to have been recognized. A color rectangle of either blue or white in the upper part of the picture together with an overlapping cloud texture rectangle gives a good reason for having recognized clouds (Fig.s 106 and 107). Note that it is not necessary to call for a precise overlap of color, texture, and shape: as is well-known for example from aquarelles, contours and colors need not fit exactly and still allow us to determine clearly what is shown.

[86] In the later versions of the system, the relative positioning was abandoned; the more precise grid positions of the rectangles are used instead. Furthermore a "density parameter" was introduced mirroring the proportion between the grid elements covered by the rectangle that have the corresponding feature and those that have not.

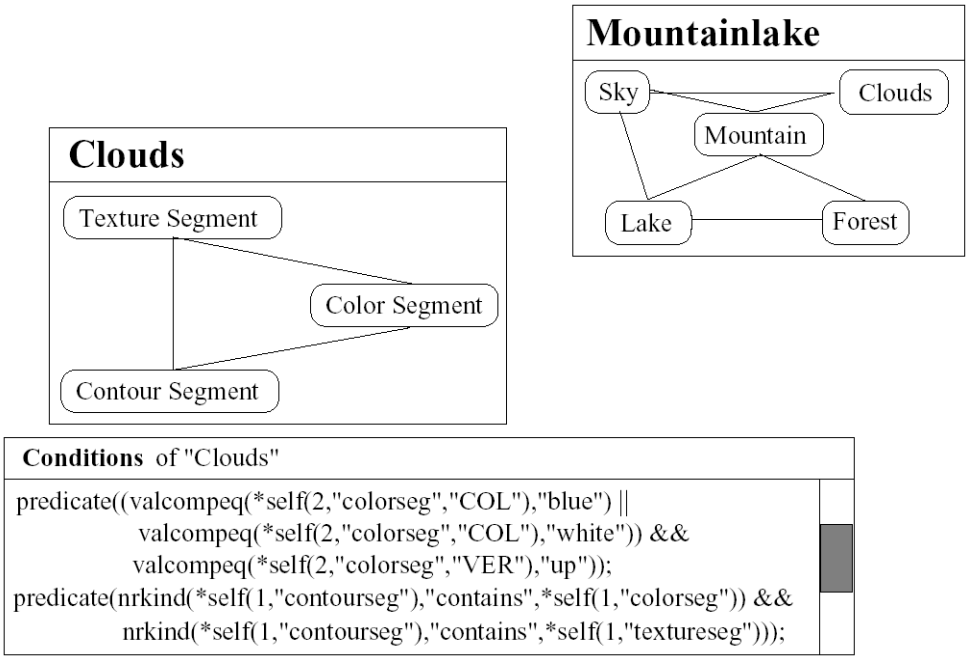


Figure 106: A Simple Object Schema ("Clouds") and a Complex Object Schema ("Mountainlake")

5.1.2 Results and Queries

The overall result of the parsing is a topological graph of primitive objects. Spatial relations over and above simple topological relations are not yet used in this version of IRIS. The resulting graph is parsed by a second graph grammar dealing expectation-driven with part-whole relations for more complex object concepts, the definition of which is encoded in the thesaurus management system TM/2. It even allows the system to classify objects that are only partially pictured. The thesaurus managing system forms the knowledge base providing the part-of relations inherent to object schemata of a certain domain. The most complex "object" types are the scene categories, like landscape, architecture photography or technical drawing that are explicitly given when a picture is to be integrated. Stating explicitly the category of a picture when adding it to the database helps significantly to determine the annotations proper, since the expectation-driven parsing can be performed in a more focused manner.[87]

The overall description of the image is finally given by one or several resulting structures reflecting the topic (e.g., mountain landscape), its particular complex constituents (e.g., snowy mountain, meadow, lake), their elements (snow, water), and the corresponding marker values, which is finally fed into the database. That is, a structured document containing not only the final interpretation but all the intermediate descriptions of the image, as well, is indexed in a text retrieval system; a user, thus, may use both syntactic and semantic descriptions for searching images with the system IRIS.

Queries to the database can be formulated on any level or combination of levels contained in the image annotation (Fig 108). Specific interfaces have been provided to ease the user specifying parameters on the lower levels. Color, for example, can be specified either by using text (a partially instantiated color rectangle description), an example

[87] Specifying the category in advance is already necessary for using an appropriate set of parameters for feature extraction.

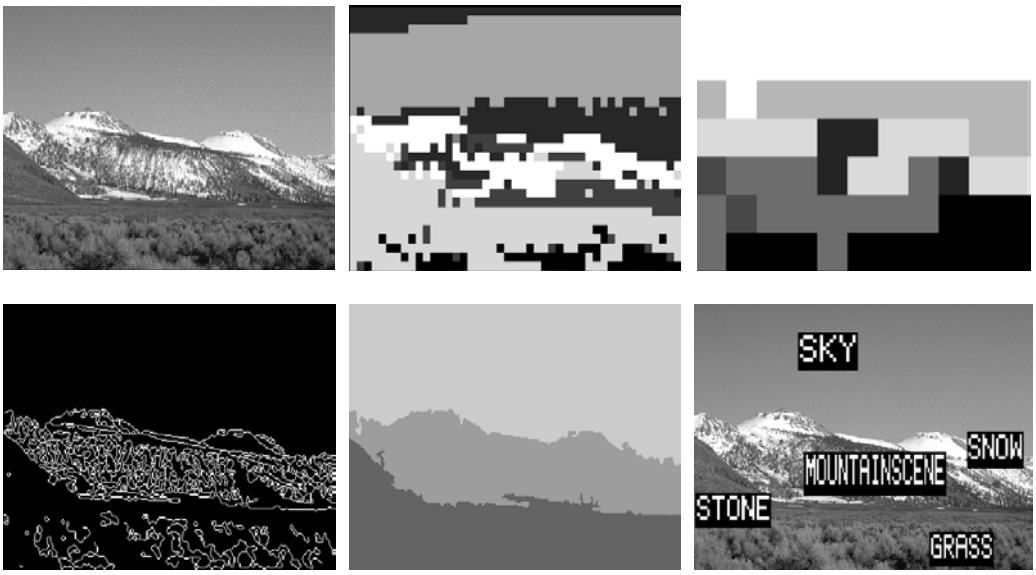


Figure 107: Visualizations of the Intermediate Analysis Results of IRIS

upper line: original image, color rectangles, texture rectangles;
lower line: contour elements, segmented regions, object association

("picked" from a given picture) or a color editor. Similarly, texture can be specified verbally by the texture category (in a partially instantiated texture rectangle description) or by an example area from a given picture. Weighed correlation measures are used for computing similarity between two feature vectors. Of course, the most complex type of request can only be stated verbally by naming the object concepts included in the scene, for example, asking for a picture with "mountain", "snow", and "lake"; or, on the most general level, by simply asking for the type of scene – "mountain scene".

Pragmatic restrictions, in particular concerning user modeling, have not been considered in IRIS. However, the image analysis has been explicitly designed in a relatively simple way so that users can more easily understand the categories used for indexing and are not mislead to ask too much "understanding" from the system. Of course, IRIS does not really understand the pictures it analyses; it is able to deliver a coarse approximation to »picture content« based on a simplified picture syntax. This approximation is proposed as a simple-to-use semi-semantic specification for a kind of image retrieval close to common-sense picture understanding.

Beside its use in picture archiving in the strict sense, the automatically derived descriptions can be integrated as special digital watermarks in pictures that are to be published in the WorldWideWeb: search engines could use that information to find more easily the pictures a user wants to find, and to block others irrelevant (or prohibited for a certain user group).

5.2 Rhetorically Enriched Pictures

The expression 'rhetorically enriched picture' has been introduced in section 4.4.1.3: Having observed that the variation of presentation styles in a picture suggests a certain figure-ground distinction over and above the mere medial spatial configuration, we have concluded that they can be easily understood as bearing a preferred reading with respect to nomination and predication. While a neutral context builder leaves the figure-ground differentiation to a complementing verbal commentary (or completely to the viewers

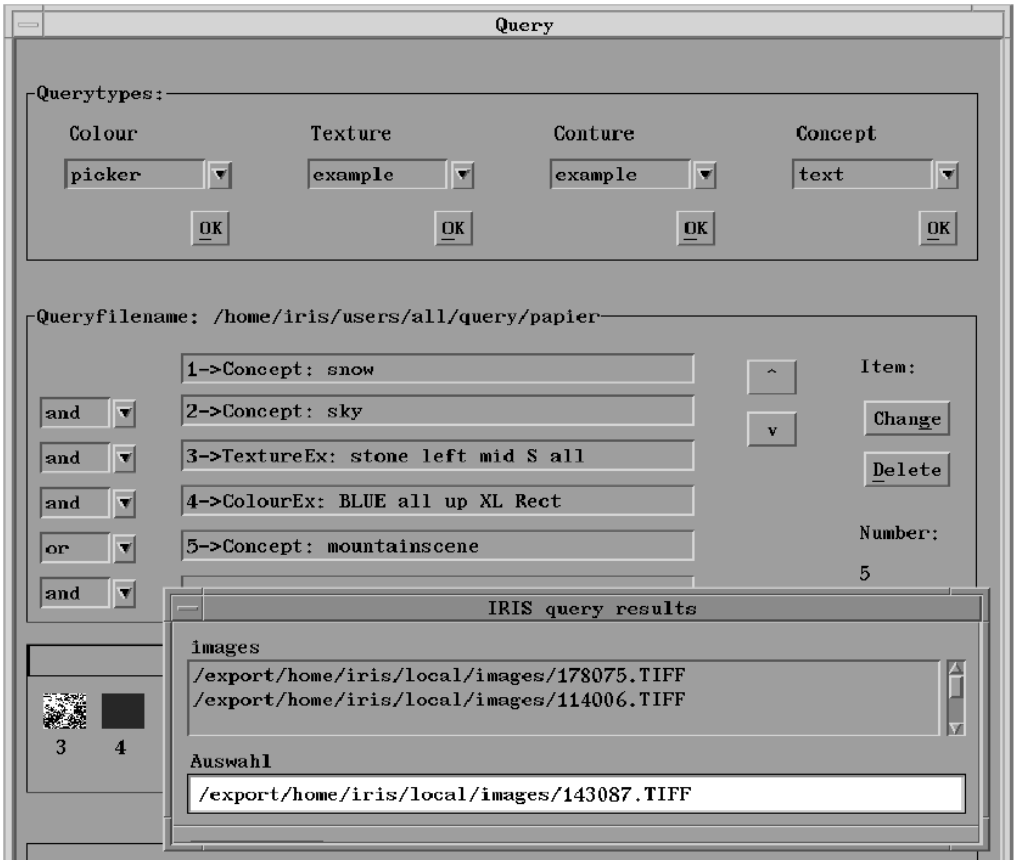


Figure 108: Query Menue of the System IRIS

and their particular interests), pictures with articulate style differences induce a specific interpretation – at least to some degree. Textbooks for anatomy, for example, avoid traditionally photographs not only because it is too complicated to erase irritating details: essentially, there is too little stylistic variation available with photography's extreme naturalism. A long tradition of artful anatomical drawing allows the designer for pictures with a better mixture of representation styles fitting their rhetoric demands.

In this case study, we are basically interested in the influence the variation of the degree of naturalism has on the (relative) rhetoric function of a pixeme. This also answers corresponding questions in image generation, for example: Which parts of the geometry in question should be presented in a naturalistic way, and why is the rest to be presented in a more "abstract" manner?

Recall in this context the meaning of the expressions 'realism' and 'naturalism' introduced before: 'realism' is the property of a representation of giving the impression of a configuration of spatial objects that is or could be found in the world. 'Naturalism' refers to the degree of a pictorial representation to which it evokes a visual impression as close as possible to that of the scene depicted. While realism is a binary category, naturalism only defines one pole of a continuous scale. Compare the two pictures in Figure 109: both represent a spatial scene in representation styles with a low degree of naturalism. What is the different meaning they suggest by means of their stylistic difference? It obviously is related to the differentiation between aspects (shown as) already known to the beholder, and new (informative) parts. The former can easily be used as anchor points for the other elements supposedly not yet known to the beholder.

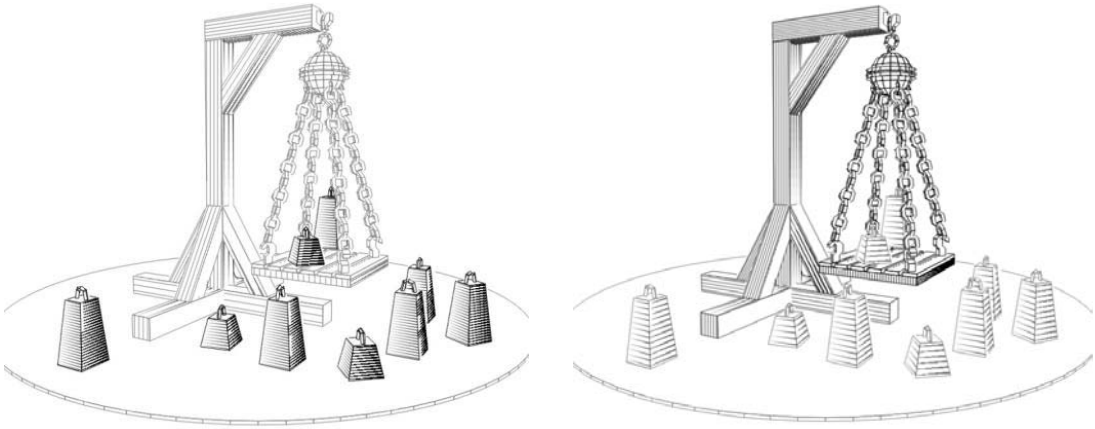


Figure 109: Two Contrasting Examples of Rhetorically Enriched Pictures

Recall also at this place the sub functions of propositional communication (cf. section 3.4.1): by nomination those aspects of an utterance are meant that refer to something already mutually known (e.g., previously mentioned). They are used to provide the other interlocutors with an anchor point for information that is new – the predication. The given anchor and the new distinction can indeed be conceived of as prime functions of rational communication in general. We therefore conceive of them as important rhetoric functions for rhetorically enriched pictures, as well, over and above the more fundamental function as context builder of pictures in general.

The various styles in pictures with mixed representation are only indirectly associated to pixemes. They relate directly to semantic elements, i.e., elements of the picture content. Let us call those parts of the underlying geometric model the 'representational elements' of the picture. The part-whole relations inherent to sortal objects organize these elements.

5.2.1 Descibing Style Parameters

In order to select style parameters depending on the rhetoric functions of nomination and predication for an example system, we first have to provide a formal language describing the style variations at hand. As we cannot deal here with the full range of syntactic and semantic factors that may add to the naturalism of a realistic picture, we restrict ourselves to a reduced and very simplified list of visual components (Table 4): First, there is color: a representational element of a picture may be ordinarily colored, uncolored or colored in an unnatural manner (e.g., duo-toned). Second, representational elements may show texture: the ordinary distribution, no texture at all, or a wrong texture (e.g., cross-hedged). Third, picture elements have form: as a respect for similarity, this dimension ranges from photo-realistically shaded projection of the full 3-D form through sketch with outlines and inner contours for indicating part-whole relations to the pure outline. Finally, the relative place of the representational element and the configuration are relevant: they may be considered as either the "natural" one or any other. The dimensions are not completely independent. Configuration is closely linked with the referents' part-whole relations, and controls how the corresponding representational elements of the parts form a representational whole. A representational element without parts has no configuration: if a representational element shows only

Table 4	Color	Texture	Form	Place	Configuration
	natural	*natural*	*shaded*	*natural*	*natural*
	uncolored	*untextured*	*inner contour*	*unnatural*	*unnatural*
	unnatural	*unnatural*	*outline*		*atomic*

outlines, all its parts are suppressed – the element becomes "atomic". In that case, the parameters of color and texture have to be adapted, too.

Any presentation style available for tele-rendering can be attributed such a style description. The "classical" photo-realistic rendering for example corresponds to *«natural color, natural texture, shaded projection, natural place, natural configuration»*, a copper plate engraving of an exploded view of a technical device to *«uncolored, unnatural texture, inner contours, unnatural place, unnatural configuration»*. The general idea is, that an interactive system uses such a characterization in its active beholder model in order to determine how a picture is to be generated for a particular user, or in its passive beholder model for evaluating a given picture.

In a picture with mixed styles, each representation element has, obviously, its own stylistic characterization: a tree of style descriptions according to the part-whole relations between the elements has to be considered. Due to the dependencies, the attribution of one value in the hierarchy puts constraints on the other values.

5.2.2 The Heuristics of Predicative Naturalism

The dual questions are then: how do we select the style parameters for a given association between representational elements and basic rhetoric functions? And: how can we reconstruct that association given the style parameter description of the representational elements of a picture? We need principles mapping one association of the hierarchy of representational elements with rhetoric basic functions onto another association of the hierarchy with style parameters, and *vice versa*.

Some presentational elements serve the communicational purpose of anchoring the place of another, usually more central element. We may assume that the beholder of that picture does not yet know that element well enough, which was the original intention to ask for that picture at all. Form and configuration of the other elements are used nominatorically, and they are sufficient for the intended user to be able to establish a context already known for the new information. All other respects of representation – color, texture, etc. – are reduced. In contrast to that, the representational element in focus is given a richer, more naturalistic appearance with more details of form and configuration, and, eventually, some color and texture.

Indeed, a representational element of a picture can take over both rhetoric functions: some of its visual properties may be nominatoric while others are predicative. This observation depends on the fact that there are no pictorial proper names, only definite descriptions. A representational element in an image is able to carry the rhetoric function of nomination merely by means of certain visual property in contrast to the other elements. Other style attributes of that element can simultaneously carry the predicative function. We therefore rather speak in the case of pictures of nominatoric and predicative properties of representational elements than of nominatoric or predicative representational elements.

As a heuristic rule derived from these arguments – a *Heuristics of Predicative Naturalism* for realistic pictures suggests itself: the parts of a spatial scene playing the role of nominatoric anchors in a picture of that scene should appear less naturalistic than the representational elements carrying the predicative properties. Correspondingly, the strange and unexpected parts are to be presented more naturalistically than the common and known.

Given a list of (visual) properties to be communicated by a picture and the list of properties that have been communicated already, the active beholder model of an interactive system has the task to select for each representational element a presentation style according to the style description. That selection must allow the system an encoding so that a subset of the nominatoric properties of that picture element (if there are any) sufficient for identification is included. Furthermore, as many of its predicative properties as possible have to be shown. While the nominatoric subset can be reduced to the minimal set satisfying a unique identification, the predicational information should be given redundantly to ensure a proper understanding. In accordance with GRICE's maxims (cf. Section 4.4.2.2), the picture is then as informative as possible (with respect to predication) and not more informative than is required (with respect to nomination).

For a certain communicative intention, this specification enables a system to choose autonomously a style that matches the situational conditions best. The Heuristics of Predicative Naturalism is to be seen as a strategic rule among others for the generation of mix-styled pictures; its results may be overwritten or modified by other strategic rules or meta-principles (e.g., consistency principles).[88]

The computational visualist can also use the heuristics for critically evaluating the rhetoric force of the interactive system's pictures by means of the passive beholder model before actually having them presented. According to the Heuristics of Predicative Naturalism, somebody receiving the picture may assume, at least as a first guess, that the producer of that picture wants him or her to understand the less naturalistic parts as nominatoric, and the more naturalistic ones as predicative. The comics examples quoted from M^{C}CLOUD in section 4.4.1.3 can be interpreted exactly in this sense.

5.2.3 Example Application of the Heuristics

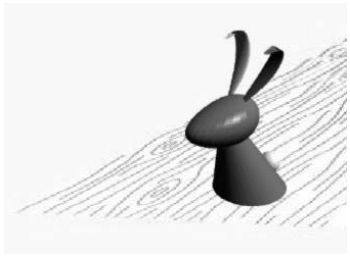
In general, the Principle of Predicative Naturalism is realized as a constraint system propagating restrictions through the hierarchy of representational elements and their visual components until a stable association has been established. The following parameters act as additional constraints:

(a) the order of degree of naturalism between the style values (e.g., the degree of naturalism decreases from 'natural color' to 'uncolored' to 'unnatural colors'),

(b) the dependencies between the different visual components considered by the style description (color, texture, form, place, configuration), and

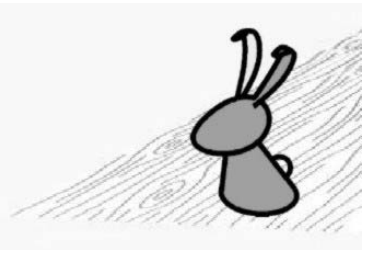
(c) the impact a particular attribution of a rhetoric function to a visual component has on the other components of that representational element.

[88] Such a set has been suggested, for example, by RIST [*1996, Sect. 5.2.2*] (cf. section 4.4.2.1).



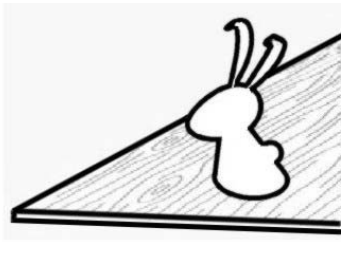
Background: all nominatoric
Bunny: place nominatoric
shape predicative
config. predicative

→ *Background: « . , . , outline, atomic»*
Bunny: « . , . , natural, natural»



Background: all nominatoric
Bunny: place nominatoric
shape nominatoric
config. predicative

→ *Background: « . , . , outline, atomic»*
Bunny: « . , . , inner contours, natural»



Backgr.: initially all nominatoric
Bunny: shape nominatoric
config. nominatoric
place predicative
(⇒ Backgr.: config. → predicative)
→ *Background: « . , . , inner contours, reduced»*
Bunny: « . , . , outline, atomic»

Figure 110: Simulated Applications of the Heuristics of Predicative Naturalism

Figure 110 presents three gray-scale examples: color and texture parameters are ignored here for simplicity.[89] Let us assume that the communicative intention given associates the background (a table top) as nominatoric, and a bunny's *shape* and the *configuration* of its (body-) parts as predicative, while the relative position of the bunny with respect to the background objects is also marked as known already to the beholders in question. Then, it is sufficient for the background to be drawn in outline and without parts ("atomic"). For the bunny however, *form* and *configuration* parameters should be maximal (i.e., value "natural"). If only the bunny's *configuration* is predicative, the *form* parameter changes to "outline": the four components are clearly discernable.

The third example is more complicated since the *place* of an element can only be predicative if the configuration of the complete scene is emphasized, as well. Therefore, using the bunny's place as predicative property has to be propagated up in the hierarchy of representation elements to the configuration of the scene, and from there down again to the place parameter of the other children, i.e., all the background elements. The inner contours of the tabletop are selected while the configuration of the bunny becomes atomic and only outlines need to be shown. The graphic clearly emphasizes the bunny being almost at the border (in consequence, its danger to fall off is here more evident).

5.3 A Border Line Case: Immersion

Images in highly immersive systems are the interactive pendant of *trompe l'œil* pictures. Using them may be seen as a borderline case of image use due to the dominance of the deceptive reception mode. As was already described for the immersant in the art work *Osmose* in section 4.4.5.3, the sign character of the presentation disappears at all: the perceptoid context builder becomes a genuin situational context. With respect to aesthetic considerations, GRAU explains in his essay on *Osmose* [*2003*]:

> *In virtual environments, a fragile, core element of art comes under threat: the observer's act of distancing that is a prerequisite for any critical reflection. Aesthetic distance always comprises the possibility of attaining an overall view, of understanding organization, structure, and function, and achieving a critical appraisal.*

[89] The system described in [HALPER *ET AL. 2002*] was used to construct parts of Figure 109. Due to its modular organisation of simple style-varying operations, it is an ideal candidate for interpreting the style descriptions finally associated with the hierarchy of representation elements of the picture.

The act of distancing is indeed not restricted to aesthetic considerations in the close sense but forms the core aspect of the symbolic mode: evoking contexts that are not the situational context at hand is the major function of signs. 'Evoking a context' means not only immediately activating certain spontaneous reactions, but also the ability of postponing those reactions to the context evoked though not present (cf. Sect. 3.5.1). Thus, one of the components defining the concept »picture« is missing. If the use of pictures in highly immersive systems aims at a reception in the pure deceptive mode by suppressing any factors able to activate the viewer's symbolic mode, then there are indeed no pictures used at all, at least for the "immersants" who do not spontaneously reflect about their situation.

Nevertheless, an important part of computational visualistics deals exactly with this borderline case of image use. As has been mentioned already in chapter 3, the producers of *trompe l'œil* pictures as well as of immersive systems cannot share the reduced reception mode: they in fact deal with pictures as perceptoid signs that are *intended* to be mistaken – at least for a time or to a certain degree. Usually, we do not meet immersants that do not at all reflect their specific situation: highly immersive systems still need so much technical effort and preparation that nobody simply find themselves in such a system without noticing it. Correspondingly, they are not in a pure deceptive mode but in the immersive mode where the postponing of spontaneous reactions is strongly reduced, not totally absent. Recall the situation in cinema when we appear to be confronted with a life-like *Tyrannosaurus rex* threatening to attack us: we know that it is an illusion (a sign we show to ourselves), consequently having a lot of spontaneous reactions that are generally postponed; but we allow those reactions to surface to some degree, which is one of the pleasures of viewing such films.

Long before expressions like 'virtual reality' or 'immersive systems' have become popular, S. LEM investigated different levels of immersion under the name of 'phantomatics' [LEM *1964*]. Present attempts are based essentially on technical devices projecting pictures on more or less flat smooth surfaces covering all of the field of view: they can be viewed in the ordinary sense.[90] Sound is being emitted by speakers in more or less sophisticated manners, also to be heard in the traditional way. Quite obviously, only distance senses can be easily deceived by that form of immersive systems. Many contact senses are much harder to be deceived – recall all the sensations from our skins. The feedback from the immersant's actions is also mostly restricted to very specific and very small channels: a mouse, a data glove, or a data suit at most.

As an improvement of immersion, LEM imagines what he calls 'peripheral phantomatics', where technical devices directly manipulate the immersant's peripheral nervous system, feeding the sensoric nerves and taking signals from the effectoric nerves: the pictures or sounds used can then be inspected only by additional devices. Still, the immersant's body with its movements and the monitoring proprioceptors form a source of "disturbance" to the immersion. Therefore, another step is introduced: LEM's expression 'central phantomatics' refers to the hypothetical technology that allows technical devices to directly manipulate the immersant's central nervous system, overwriting any bodily signals including those from the sense organs, and intercepting any nerve pulses controlling (real) motion. That is: for the immersant, the physical body is completely replaced by the avatar. He or she seemingly exists only in "cyberspace".

[90] Current experimental devices projecting light directly into the eye form the only exception known to the author so far.

Beside the aspect of plausibility, LEM is basically interested in the epistemological question whether (and how) an immersant of any of those levels would be able to detect the deception – indeed a modern pendant on DESCARTES' reflection on the nature of truth under the assumption of a deceiving deity.[91] This is not the place to investigate those problems.

Of course: the expression 'immersion' is not only a question of technology – more likely a question of purpose and concentration. Humans are very well able to be completely absorbed in reading a novel ignoring most of their situational context (even up to the bodily needs to some degree) evoking mental imaginations of what they read. High degrees of technically induced deceptive mode are relevant in simulations, e.g. for pilot training. In other applications, the equilibrium between the symbolic and deceptive components of reception is more complicated. Different purposes lead essentially to different possibilities of interactions. Two particular cases – virtual architecture and virtual institutes – are presented in the following together with a description of the particular conditions and intentions of application, demonstrating alternative needs for deceptive or immersive reception modes of the pictures used.

5.3.1 Virtual Architectur: The Atmosphere Projekt

The first example is the computational reconstruction of an important *Jugendstil* house built by PETER BEHRENS (1868–1940) in Darmstadt, which is not preserved in its original state. The house was designed as a piece of the exhibition *Ein Dokument deutscher Kunst* that was prepared by the artist colony of Darmstadt in 1901. It represents a unique *Gesamtkunstwerk*: Apart from the architecture of the façade and the exterior disposition, all details of the interior decorations are based on designs of BEHRENS himself. This does not only include the doors, windows, carpets and wallpapers. The form of furniture, lamps, glasses, chinas, and cutlery, even music instruments, inkstands, and jewelery fitting to the house's aesthetic conception are based on inventions of BEHRENS. In this unique fashioning of human living space characteristic for Art Nouveau, a specific conception of life was expressed. It is the approach of using artistic abilities to transform the environment in a beautiful and reasonable manner in order to harmonize again humaneness and the technical development [BUCHHOLZ *2000*; BEHRENS *1901, 3–6*]. During the 19th century, the two aspects had developed more and more into quite different directions in Europe, a disintegration of the manners of life that pushed forcefully into public conscience at the end of that century and gave rise to a large number of reform-oriented approaches, which to our days have a strong influence.[92] The functional and aesthetic investigation of an integral ensemble like the House Behrens is an important building block toward a proper understanding of the characteristic conception of life of that time and its lingering influence on us.

[91] While LEM keeps an ironic distance to such technologies and is interested mainly in epistemological and ethical questions, other scientists are not so prudent: recall, for example, M. MINSKY's public fantasy about all humans being "equipped" with an implanted computer interface to their brains. The creativity of narrative artists had been excited by such scenarios, as well; beside FASSBINDER's TV production *Welt am Draht* [*1973*], CRONENBERG's film *eXistenCe* [*1999*] is one of the more interesting results.

[92] Just as a few examples: fitness studios and ecological agriculture, functional architecture and feminism, health food and artificial tanning; they all can easily be traced back to the broad movement of "Lebensreform" about a century ago [BUCHHOLZ *ET AL. 2001*].

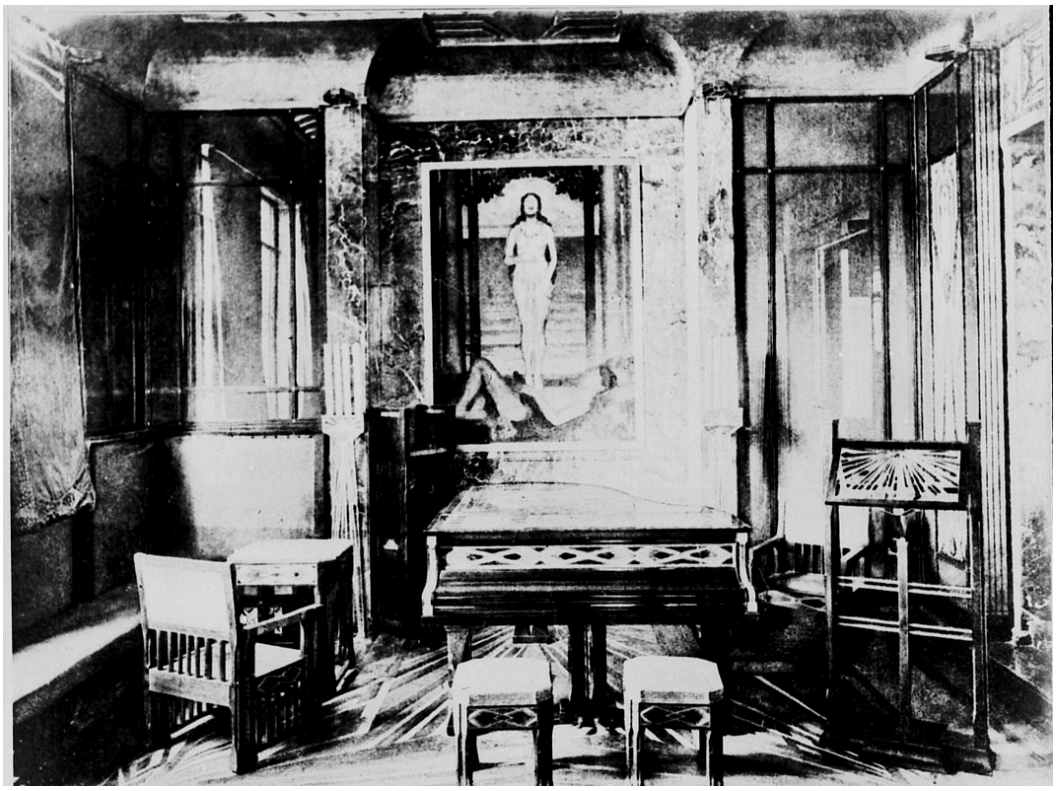


Figure 111: Contemporary Photography of the Music Room. The door to the dining-room is just outside the right frame border

However, the house was destroyed severely during Word War II. The façade was reconstructed afterwards with only minor alterations, but the interior decoration is utterly lost.[93] The original partitioning of rooms has not been restored; all decorative elements are destroyed. A reconstruction by means of computational visualistics – a virtual architecture – must, then, appear as a very plausible approach [FORTE & SILIOTTI *1997*].

Such a virtual reconstruction has to be based on data about the original contexts as precise as possible. In the case of BEHRENS' house in Darmstadt, there is a sufficient amount of details available at least for the two central rooms of the ground level: the fact that pieces of Art Nouveau have been broadly documented in illustrated papers of that time does not only demonstrate how important the underlying conception of life was rated then; it obviously is quite helpful for the virtual reconstruction, as well. The journal *Deutsche Kunst und Dekoration* appearing in Darmstadt published an extensive article about BEHRENS' house with floor plans, sketches, and many large black-and-white photographs [BEHRENS & BREYSIG *1901/02*]. A similar paper appeared in the journal *Dekorative Kunst* [SCHEFFLER *1902*]. A special catalogue was produced for the exhibition covering exclusively BEHRENS' house: it contains an introduction by BEHRENS and a list of all the enterprises that constructed the objects designed [BEHRENS *1901*]. Finally, numerous modern color photographs are available of those parts of the furniture that have survived World War II and are exhibited in several museums (e.g., [INSTITUT MATHILDENHÖHE DARMSTADT w/o *y.,* 6–27].

The two rooms in our focus of attention have been discussed extensively in contemporary architecture critiques: they are the music room and the dining room just beside

[93] Furniture and other movable parts had been already removed from the house since BEHRENS moved to Düsseldorf in 1903. Fortunately, some of the furniture was therefore spared from the destruction.

Figure 112: Contemporary Photography of the Dining-Room. The door to the music room is partially visible at the left side

each other (Fig's 111 & 112). Both rooms serve social purposes and are connected by a large double door with partitioned wings. The music room is higher in order to evoke a particular atmospheric effect described by BEHRENS [*1901, 8 f.*]:

> *In order to heighten the music room in accordance with its purpose with respect to the rooms around it – the true festive room of the house – it was necessary to lead down two steps from the hall and simultaneously to lift the ceiling approximately for the same distance compared to the adjacent dining room. The two steps have the practical purpose named; but also the other, spiritualized one: to lend a rhythmical movement to the traffic between dining room and music room. Stepping down gives us the feeling of being prepared for something; stepping up evokes the one of lifting to something. And in those feelings, very essential moods of humanity can be recognized.*

Numerous details and materials have to be considered: The floor of the music room consists of a parquet with seven different woods forming a linear geometric pattern. The dining-room has a floor of mosaic in a curved pattern. The steps leading from music room to dining-room are of a pink marble, the music room's walls are covered with grey marble and blue reflecting glas. Above the door to the dining room and at both sides, additional mosaics are placed. The walls of the dining-room are interrupted by large windows above silver-coated heating grills. Between some of the windows, crystal mirrors are installed. Other parts of the walls are panelled in white-lacquered wood below a frieze of damast. The ceilings of the two rooms are richly decorated: in the music room, it consists of gold-painted wood with another linear ornament; the dining-room's ceiling has curved stucco ornaments with some of the intermediate spaces coated with silver. The doors are also particularly elaborated. The door from the music

Figure 113: Screenshot from the Virtual House Behrens: View from the Dining-Room into the Music Room (with door to the the hall)

room to the hall is adorned with a linear ornament in golden aluminium bronze similar to the music room ceiling. The door between music room and dining-room is white on the dining-room side with a simple curved ornament; toward the music room it shows "an uninterrupted surface of noble silver-maple without a single ornamental line" [BEHRENS & BREYSIG *1901/02, 148*].

The interior decorations of the House Behrens vary a small number of different ornaments. The decorative design of the dining room is developed from the basic figure of two intersected curves. They appear in the simple form in the gratings of the cupboard glasses and in the mirror cuttings. A more complex derivation is shown in the heating grills, where different degrees of sinuosity are combined. The curved line induces verve and movement into the dining room. In contrast, all forms in the music room are developed from a linear base, a rhomb, which develops into the complicated form of a crystal. Decoration and furniture of the music room induce the impression of static calmness.

Taking into account the conceptual background of *Jugendstil* design mentioned above and the particular emphasis BEHRENS put on designing a complex but uniquely coherent whole, any alteration of a detail in the virtual reconstruction can destroy the intended use. Unfortunately, many of the colors and texture are uncertain. Black and white photographs only hint at the relative luminescence. Verbal descriptions of the colors are often rather exuberant but obviously of limited help for the computational visualist.

The success of a project like the reconstruction of the House Behrens by means of computational visualistics depends, thus, on an intense cooperation between computational visualist and art scientist. Decisions have to be made concerning the colors or textures to be used. Has a given texture to be modified? In which way? What are the criteria? Is it technically possible? Answers depend essentially on the precise purpose of the immersive images to be produced, and in particular on the addressee. Detecting the authenticity of colors and texture falls, of course, essentially in the domain of the art scien-

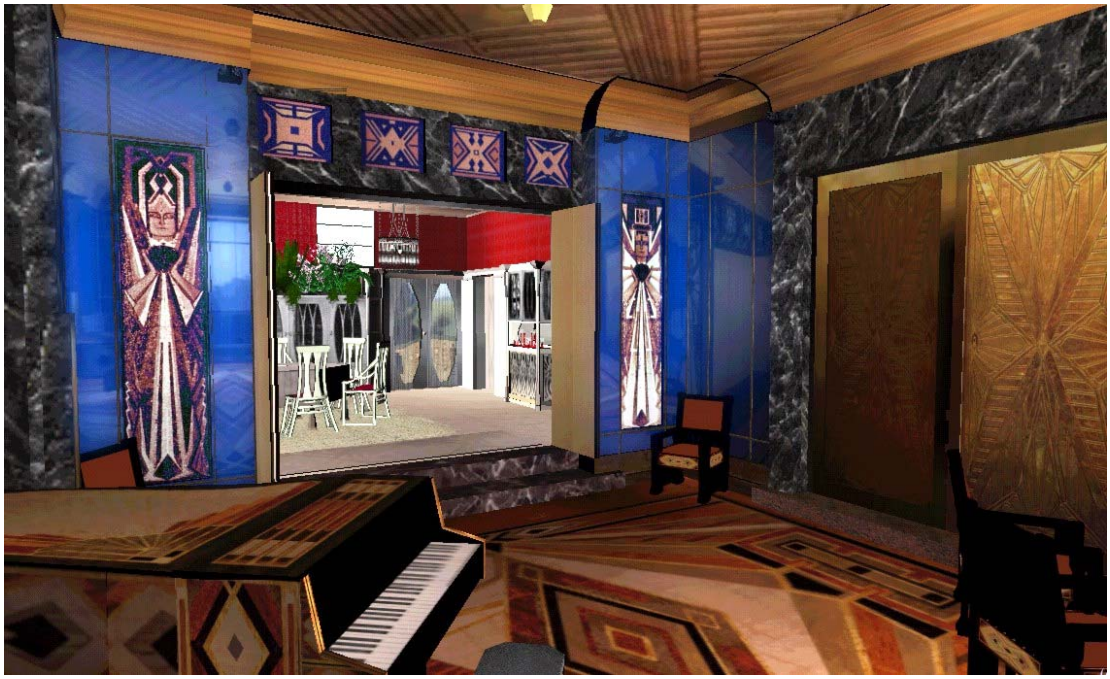


Figure 114: View of the Reconstructed Music Room toward the Dining Room

tist; but "trial shots" generated by the computational visualist are certainly quite helpful even if they are still far from an acceptable end result. In a way, the visual intermediate results (e.g., (Fig's 113 & 114) serve as a virtual experiment in which the effects of decisions about colors and textures can be concretely studied. Thus, the sign character explicitly controls the deception.

The border of present immersive technology can be clearly demonstrated when we consider the broad use of reflective surfaces in the two rooms under investigation: that a material is highly reflective is, of course, not a major problem for computational visualistics: but what happens if the viewer looks at one of the blue glasses in the music room from an angle of approximately 90°? In the original (not virtual) room, she could then see her own mirror image – an effect BEHRENS certainly has taken into account when planning the house. It is, of course, impossible to model for every user of an interactive system that provides virtual strolls through the reconstructed House Behrens a fitting avatar that could be seen in the virtual mirrors. The geometrical and optical specifications of the body of that user in its present bearing would have to be included in the model. On the other hand: being invisible in the virtual context or having no body is certainly not a satisfying solution, too. Finally, a "digital dummy" could appear in the virtual mirror representing the user, though it is probable that a user may mistake it not for the own reflection but for the avatar of another user. A solution is not yet palpable.

In the perspective of pictorial pragmatics, the problem offers an interesting aspect apart from aesthetical considerations: How does the missing (or wrong) reflection disturb the intention of the reconstruction of a lost *Gesamtkunstwerk*: i.e., to maintain the deceptive mode of reception? In general, the virtual reconstruction of the House Behrens has to investigate the problem of how much intentional deviation from the original is possible without upsetting the integral atmosphere of the environment. On the other side: how much deviation is necessary in order to mark clearly those parts that are not (or not certainly) integrated in their original appearance?

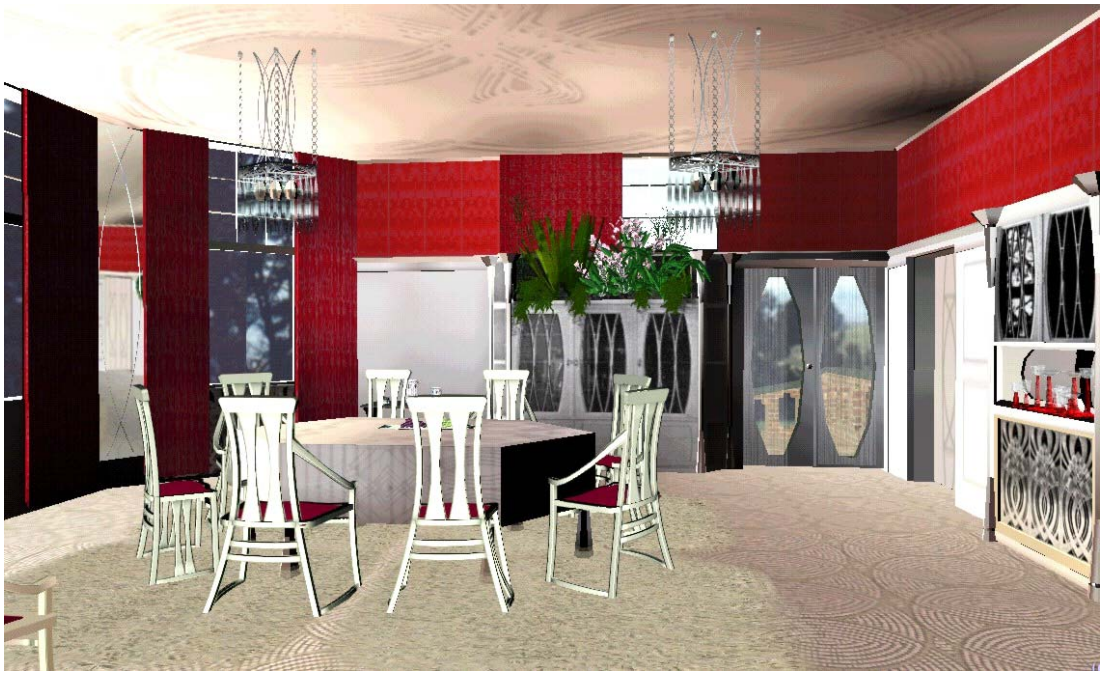


Figure 115: Two Screenshots from Approximately the Same Perspective – Daylight Atmosphere …

No doubt: virtual architecture can provide experiences over and above those of contemporary photographs: for example, views from arbitrary perspectives, and – up to a degree – the atmospheric synthesis of the colors and materials used. The effects of different lighting provide another prominent example (compare the two parts of Fig. 115). Mediating further atmospheric aspects in an adequate manner is more complicated: the experience of stepping up or down between music room and dining-room, for example, can hardly be gained in front of a computer screen with the mouse as means of navigation. The same holds true for the authentical impression of the size of the rooms and the objects within: in order to "really" walk though the computationally reconstructed house, we have to employ a highly immersive system, like a CAVE [CRUZ-NEIRA *ET AL. 1992*] – a cubic room with stereo projections on (almost) all walls on which pictures in the correct perspective relative to the position of the user in the room are projected. The position and view direction of the user are registered; they influence in real-time the projections. Shutter glasses let the beholders see objects in stereovision. Similarly, corresponding sounds – e.g., of steps, closing doors, or drawn curtains – adapted to the actual position of the immersant with a surround sound system enhance the deceptive feeling of being present in that virtual reality.

The computational effort for the different projections of the House Behrens simultaneously necessary in sufficient detail is still rather too high. Apart from that, using a CAVE for presentation certainly gives a better basis for studying atmospheric aspects of such an artistic ensemble. At least, the proportions of the rooms or their acoustics can be perceived in an adequate manner, quite close to the original. But even for that form of presentation of a virtual architecture, we can easily find problematic aspects of atmospheric effects. Although the visual impression paired with corresponding stepping sounds suggest that we are stepping up from the music room to the dining room, we do not use the muscles of our legs in the same way as in real life.

... vs. Nocturnal Atmosphere (and a few more alterations, e.g. drawn curtains)

With the questions of aesthetical atmosphere, we meet, it seems, a fundamental difficulty of virtual architecture and its use of computer-generated images. The problem is connected with BEHRENS' aim of artistically permeating *all* aspects of life in the house. In the House Behrens, *everything* is designed towards one unique homogeneous effect. That impression of aesthetical homogeneity has been emphasized by the contemporary critiques (cf., e.g., [MEIER-GRAEFE *1901, 482–484*] or [SCHÄFER *1901, 39*]). In order to adequately reproduce that impression, it is however necessary to rather use replicas instead of images since the specific distance that always separates the beholders from the image referents due to the sign character of the picture easily may "poison" the atmosphere and disturb the integral impression intended.

However, it is not the task of virtual architecture to provide room that can be (virtually) inhabited: We are interested in gaining a sensible impression of the house that is sufficient to understand how the abstractly described desire of BEHRENS (and other artists of Art Nouveau) – to extensively embellish everyday life – is put into concrete effect. It is quite unclear, how close we have to come to the original to gain a "sufficient" impression of the characteristic atmosphere.

5.3.2 Types of Use of Virtual Architecture

The degree to which the original atmosphere has to be evoked varies with the different aims of such a virtual architectural reconstruction. There are essentially three goals determining the design: the presentation should be mainly (1) educational, (2) scientific or (3) entertaining.

Among the educational goals, the focus can either be on the mediation of the historical appearance of *one* individual object, which together with other exhibits and apart from the virtual architecture is expected to lead to corresponding new knowledge of the beholders/users: "That's how that object did look like, that approximately was the integral effect." Quite easily, we can imagine to have a guided tour in a museum through

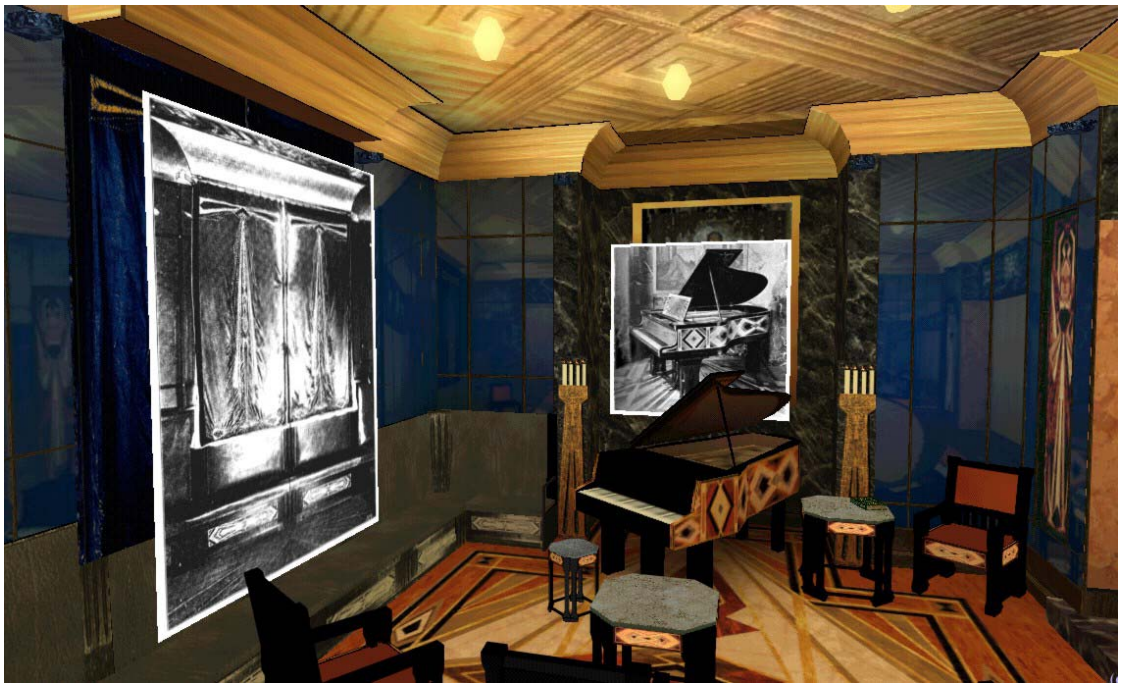
Figure 116: Another View from the Music Room – Standard Presentation and …

the virtual House Behrens in a CAVE-like presentation engine; or to have the opportunity at a smaller screen to stroll on our own (with a mouse) discovering the atmospheres of the reconstruction. But the presentation can also be employed for demonstrating *general* aspects of Art Nouveau or the special idiomatics of BEHRENS' early style. In that case, it is reasonable to directly integrate other exhibit in the CAVE presentation: the general stream behind the individual aspects of the house would not become clear otherwise.

An example is shown in the two parts of Figure 116: the visitors do usually not know how the original did look like or what have been the sources for the reconstruction. Integrating on demand posters of contemporary photographs of the original – or of pictures of other important buildings of Art Nouveau, its precursors and successors; or of further documents of relevance – into the reconstructed architecture gives the users an opportunity to grasp corresponding abstract aspects. In a way, the rooms of the house then become the place of a secondary exhibition.

Apart from those "objective" variants, the peculiar problems of virtual architecture can also be the theme of such a presentation: the visitors to the museum are told on the meta-communicative level where the computer-generated pictures are authentical, and where more or less plausible decisions of the developers had to substitute the unknown details of the original (together with their reasons). In that case, too, it is necessary to integrate more information into the images over and above the architecture as such in order to reach the communicative goal intended.

In the framework of educational uses, the focus of interest can finally be on showing how such a virtual world for a presentation at a museum is created at all. In that case, the reconstructed House Behrens is just a more or less arbitrary example. In general for educational uses, the option of moving freely in the virtual rooms is secondary compared to the necessity of showing pictures or pre-rendered animations with as much detail as possible in a most naturalistic manner.



… with Embedded Sources: Contemporary Photographs Shown and Commented on Demand

Quite different preferences follow if we consider an application scenario with art scientists using the CAVE presentation to aid the discussion in their discipline. Art scientists assume that BEHRENS' conception of his house in Darmstadt followed a general aesthetic plan: by means of the homogenous design, the architecture must have, he believed, a positive influence on all aspects of the life of its inhabitants. If that assumption really plays a role in the discussions of art scientists, then mere debates must remain pale if not sterile if the effects on life can only be investigated in blind theory. As we have seen, the subtle aspects of aesthetic atmosphere in particular depend on concrete experience. Thus, the CAVE presentation allows the art scientists to study at a vivid example the consequences of certain theses in art science. As the atmospheric impressiveness of the interior design of the House Behrens results on principles of form that are systematically applied, it is a central task for art science to propose candidates for such principles and elaborate the borders of their effects. With the virtual reconstruction, the consequences of such principles can be accessed in a concrete form for comparison. The central task of the computational visualist besides providing the immersive system as such is, then, to develop tools for the art scientist to easily control the degree of deception or symbolic distance by changing between different versions of the model (or by directly modifying it).

Take for example the parquet floor of the music room with its complicated design in crystal shape. All we have is a set of black and white photographs showing parts of the floor in perspective distortion, and a list of the seven types of woods used. Fortunately, the parts depicted in the photographs are sufficient to reconstruct the complete geometrical pattern of the parquet (Fig. 117 left side). However, the association between the parts of the pattern and the types of wood is obvious only for one type: ebony is clearly the darkest wood employed. In order to easily check different possibilities, a "parquet editor" integrated in the virtual reconstruction is highly helpful: grouping the parts of a parquet (or carpet) pattern that are to be associated to the same texture, attributing the elements of that group with different texture parameters (size, set-off, rotation), and as-

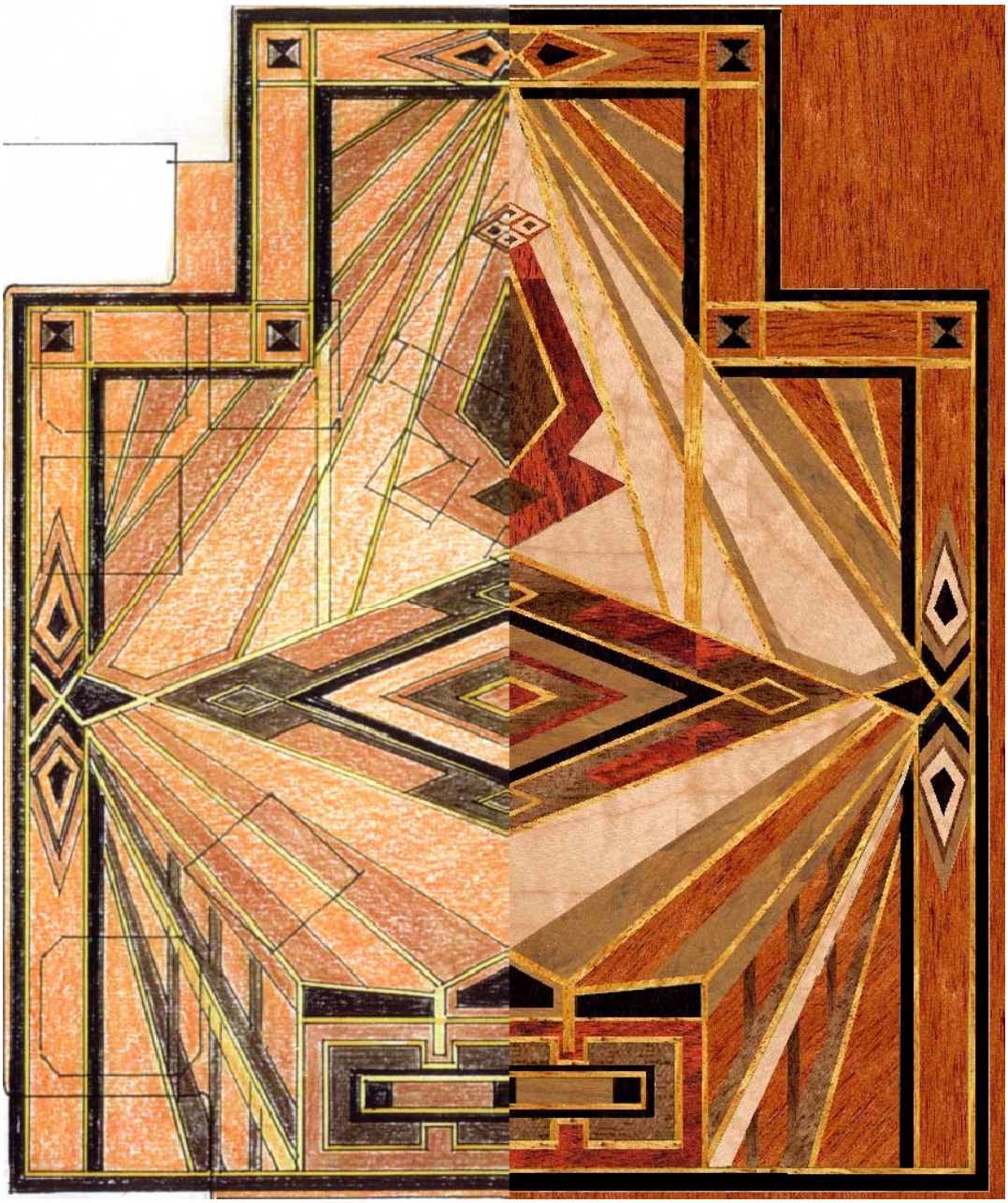


Figure 117: The Reconstructed Floor of the Music Room: Sketch and Textured Version

sociating the groups with alternative texture images are the main tasks, an art scientist might want to perform in an easily controllable manner so that the results are immediately visible in the virtual building.[94]

The virtual House Behrens could finally be set up for entertaining purposes. For example, it could be used as the scenario for a computer game. In contrast to educational or scientific uses, it is then less important to reconstruct the building as originally as possible. Major divergences are acceptable depending on the background story of the

[94] Such an editor deals only with syntactic aspects of »image«: geometric parts and textures. Used as a texture in the actual picture of the room, the result (e.g., Fig. 117 right side) does actually not appear as a picture on its own.

Figure 118: An Alternative Presentation with Sketches for Uncertain Textures

game. Interactivity is the dominating factor here followed by the degree of naturalism of the presentation. The amount of detail or alternative presentations are of little interest. More important is the option to meet other players online in the virtual building.

Due to the present state of the art of multi-user 3D engines, a corresponding reconstruction cannot yet suffice the demands for straightforward educational uses – not to mention scientific applications. In order to guarantee interactions in real-time for many users all around the world, the model has to be really simple, the textures rather reduced.[95] In principle, a homogeneously designed impression and a specific atmosphere are important, too, for game scenarios. But the aesthetics of computer games has little in common with BEHRENS' intention of beautifying everyday life by means of the house in which that life takes place, and of composing all of its elements as explicit parts of a whole. This is at least true for the present; a more elaborated investigation may however result also in hints for more ambitious computer games that take up BEHRENS' goal in another form.

* * *

Several attempts have been made to gain a clear view about the problems involved. They provided the screenshots used in this section. In particular, one approach was based on *Adobe's Atmosphere* – a program to construct and browse interactive 3D worlds in the Internet.[96] The system had the specific advan-

[95] Despite the probably significant deviations from the original, the entertaining use can nevertheless be combined with the educational setting: a computer game provided by a museum, where players learn implicitly and in an entertaining manner about, e.g., Art Nouveau, *Jugendstil*, Vienna secession, the historical context of the exhibition in Darmstadt 1901, the other exhibits and their history, and the peculiarities of BEHRENS' architecture, and, maybe, about the differences between the coarse virtual model of the game and the subtle composition of the original.

[96] The coincidence of the names of the project and the tool used has not been intended.

Figure 119: Another Perspective in the Alternative Presentation

tage to allow several users to meet in the virtual context, thus forming a good basis for modeling the House Behrens for the web presentation of a museum (Fig's 113, 114, 115 & 116). Since that system was still in its early developmental phase, it did originally not provide all the features necessary or desirable for the task. An alternative modeling based on *3D Studio Max* and Spinor's *Shark* 3D *Game Engine* allows us to use more kinds of shaders. Here, non-naturalistic textures have been tested when the original texture is unavailable (Fig's 118 & 119). Although not shown here, a simple static model in 3D Studio Max of the Behrens house was engaged at the beginning to produce a few short films in order to demonstrate to the co-operating art scientists the general possibilities and borders of computational visualistics. Another experiment was based on the level editor of the computer game *Thief: The DarkProject* [Looking Glass *1998*]: without much detail, it led to a first coarse version with real-time interaction and a plausible but mostly pre-defined integration of according sounds (stepping, doors etc.).[97]

5.3.3 The Virtual Institute of Image Science

A quite different need for immersive systems appears if we consider the idea of a virtual institute. The expression 'virtual institute' has still a rather unspecific meaning. Basically, it refers to an immersive system (in a rather broad meaning of 'immersive') accessible by means of a computer network from different places and allowing the users to perform tasks as in a non-virtual scientific institute. Virtual institutes are virtual insofar as no physical building or meeting place is required; but certainly, members of such an institute must be real persons. A virtual institute is, thus, like a virtual community, though without using the characteristics of disguise common to the latter. It may or may not adhere to the building metaphor like a virtual architecture by providing an immersive 3D platform with offices, meeting halls, foyers, galleries, and libraries. Quite different platforms for virtual institutes emphasize either the immersion aspect or

[97] Cf. also http://www.jrjs.de/Work/Projects/Behrens/

the communication aspect. The decision for a platform depends on the goals pursued with the institute: text-based chat systems allow virtual communities to flourish, single-user VRML scenes convey a highly immersive 3D impression to its users. However, both are adequate for certain tasks only. Again, it is helpful to distinguish three major application areas: research, presentation, and communicative work. The Virtual Institute for Image Science (VIB: www.bildwissenschaft.org) is almost exclusively designed for the third task [SACHS-HOMBACH & SCHIRRA *2002*]: to provide a working space persons can share for joint projects despite being physically separated.

The Virtual Institute of Image Science has been created as a platform for electronic communication in order to simplify the co-ordination of projects of a particular scientific community. The initial motivation for setting up the virtual institute was essentially the attempt to support various interdisciplinary research projects between image scientists that mostly live and work at locations far apart. The general intention was to encourage the communication between those researchers. An immersive component was not considered in the beginning. However, the first approach has not been too successful. A characteristic of present scientific authoring had its bad influence, as it hardly ever seems to happen that scientists deliver their papers before deadline. Such a "production just in time" does not really allow for an extended discussion or coordination process before publishing. Scientists (at least in the humanities) might also tend to follow primarily their idiosyncrasies and refuse to have somebody else see (not to speak of 'discuss') their papers in an unfinished state. Quite interestingly, if you meet the very same persons on a conference, and talk to them face to face, such reserve to discuss thoughts in the making is often absent: an observation that leads directly to considering 3D virtual meeting places as a better tool for the initial task of the VIB.

Nevertheless, we have been quite successful in originally bringing together scientists from the different disciplines related to image science. The motivation for the VIB has subsequently changed from coordinating single projects to establishing a novel scientific approach: the "unified image science" (or general visualistics) . In consequence, the Virtual Institute of Image Science has become the crystallization core for developing a compensation for the lacking institutional background for image scientists. This shift in the conception of the VIB opens a wide variety of useful functions to be made available for the members of the institute but also for the general public. An important step is to think about how to make the platform attractive for the members. One has to motivate them to spend time in the virtual institute, and to use the functionality it offers. On a general level, interesting functionalities for the VIB crystallize around (a) accessing data, and (b) meeting people. In other words, we need, on the one hand, a large database that contains a critical amount of relevant formal and informal information, and on the other hand, communication facilities.

The database essentially works as a library or media archive. It allows the users to easily access texts or other presentations relevant for their present interests. Adequate facilities to browse and search must be offered. Beside the user-friendliness, it is mainly a question of "critical mass" for the database (and with it the virtual institute) to become sufficiently interesting for the members of the community.

The information one can access is one reason for a scientist to spend time in a virtual institute like the VIB. But for the overall goal the facilities to meet people, to communicate, and to cooperate are equally important: the VIB also needs a virtual meeting place that is easily accessible. Here, the members have a chance to contact or meet other members and organize directly – "face to face" – research cooperation in every respect.

Table 5	immersive task	communicative task
solitary task	*Showcase for public* *Orientation and leisure* ...	*Reading / Writing papers* *Searching the database* ...
cooperative task	*Informal discussion* *Video conferencing* ...	*Formal meeting* *Passing documents* ..

Besides getting in touch with experts the meeting place should stimulate and support the starting of new projects, the organizing of conferences, and other informal "networking". The corresponding facilities must also encourage novel ways for publication, which include possibilities of reading, writing, rewriting, and reviewing texts together.

From a more general point of view, it is enlightening to sort the functions of the virtual institute in those that afford a high degree of immersion, and those that depend on dense forms of communication. Furthermore, we have to distinguish between those tasks that need the cooperation of many users interacting directly with each other from those that involve single users working solitarily at a time (cf. Table 5). Note that communication does not necessarily need direct multi-user cooperation: texts are a typical example, since they mediate communication but do not call for all participants to be present simultaneously.

Researchers in image science are trained to, and indeed have to publish text, just like other researchers. Reading and writing texts are demanding solitary tasks in the context of a virtual institute, which call for enough screen space, a familiar user interface, high responsiveness, and, last not least, a quiet, relatively undisturbed context. This is, in our eyes, the main reason why a complete switch to a 3D environment for the VIB is not desirable, as long as most users first have to struggle with usability and high demands on hard- and software. Correspondingly, most of the database functions, e.g., paper collections, are rather implemented in a "classical" form as web pages, i.e., for solitary, non-immersive uses.

Cooperatively writing a paper or research proposal depends on the solitary work of the co-authors coordinated by additional formal meetings and the exchange of documents. Web-based communication applications like file transfer, email, chat, and audio streaming are used to that purpose side-by side with classical tele-communication, like telephone and facsimile. However, "even with all those organizational and technical facilities at hand, it will still be very difficult to deal with the way of cooperation that researchers use most, namely the informal 'ad hoc encounter' that, via exchange of knowledge as well as gossip, opinions, etc., can lead to new research cooperation" [LUBICH *1995, 73*]. Usually, cooperation is clearly associated with a specific, formalized outcome: a paper, a journal, a research project. "Although it is acknowledged that formal meetings, etc. are often part of a research cooperation, the emphasis is clearly on the informal part of cooperation, whose dynamics have been much less investigated and – in comparison to strictly formal interactions – are harder to model" [LUBICH *1995, 67*] . Physical proximity evidently plays a large role in successful scientific cooperation.

If we agree on that informal ad hoc encounters play such an important role within scientific communities it is likely that a 3D environment is worth a try for supporting the informal communicative functions. After all, it mimics exactly the necessary physical proximity between interlocutors. As a first test in that direction, a 3D environment has

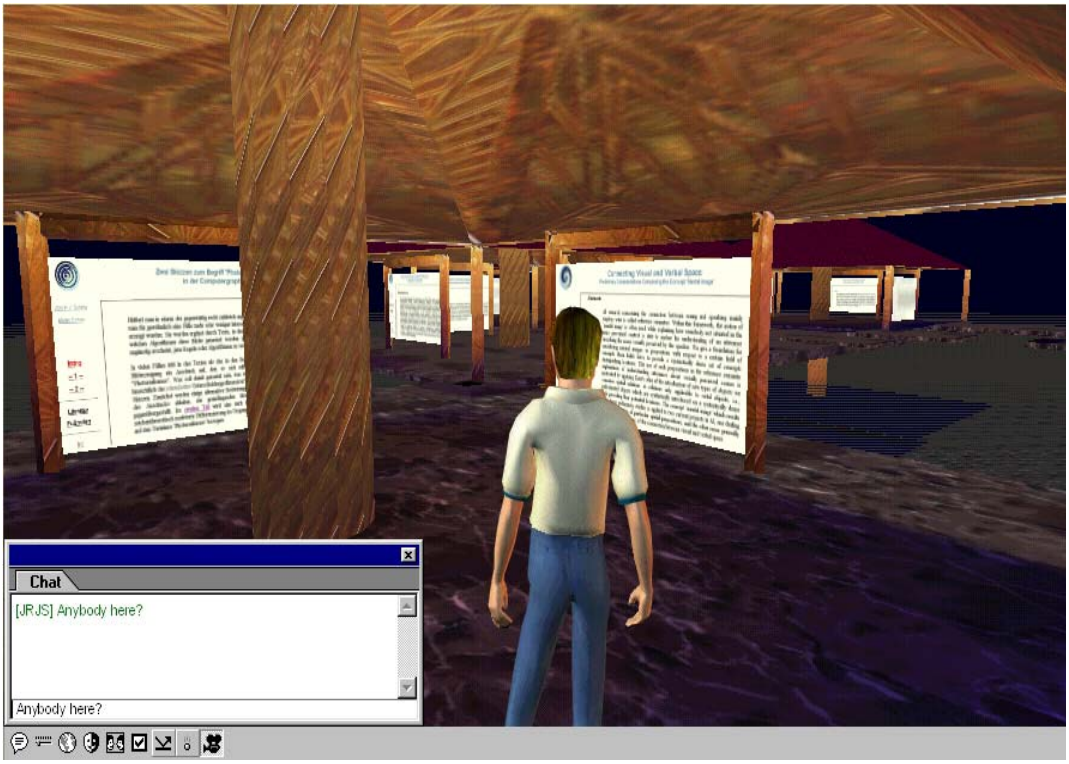


Figure 120: Experimental 3D Environment Combining Library and Meeting Place Functions

been set up to access an online paper collection simultaneously serving as a virtual meeting place (Fig. 120). The actual information of each paper is still offered as a classical *html* page. Several persons might access those pages simultaneously, too, without using the 3D environment; but they would not know from each other. When using the immersive system, simultaneous users are aware of each other, e.g., as interested in the same paper of the collection. Even more, a chat function allows them to discuss matters related to the papers in a rather informal manner (among other things they may find worth talking about).

This, at least, is the theory. One major problem with such a 3D meeting place is generally that it is often accessed by a user for relatively short times only. That is, the chance to really meet somebody else (anybody!) without having an explicit appointment is usually quite small if the site is not visited by many thousands of users a day. Here we come again across the problem of motivating people to use such functionality. Only if a "critical mass" is reached in the probability to actually meet an interesting interlocutor, the virtual reality can attract more users or more frequent accesses, and thus become a *working* meeting place at all. If we assume that a relatively closed group of users with common interests and other paths of communication are addressed, as in the case of the VIB, their effort in time and concentration used for entering the virtual meeting place must be worthwhile, or the members cease to come back and continue to use just telephone and email.

Informal meetings in a real institute may best be associated with unplanned "kitchen encounters" where one member meets accidentally another member at the coffee machine or water station, etc. A spontaneous conversation may start – leading to just those informal interactions relevant for the scientific cooperation in the institute. Similarly, tea breaks at conferences or workshops are popular not mainly for relieving thirst: they

provide exactly the opportunity for unplanned encounters from which informal meetings spark off.

Quite obviously, providing the virtual institute additionally with a virtual coffee machine will not be an adequate adaptation. The characteristics of the "kitchen encounters" are not preserved. It is essential for that (real) scenario that there is (*i*) a strong physiological need for the members to move physically to that place with (*ii*) no immediate intention to perform some work there while (*iii*) being also open for social interaction. The second criterion is important since a primary intention to do some other work might certainly reduce the opportunity to chat with a colleague met by chance. This does not change for the virtual form of the institute, nor does the third item. A plausible adjustment to the virtual institute must be found essentially for the first point. Instead of the physiological cause related to the institute's tea kitchen, a high motivation must be installed unobtrusively for the members to come to or pass through the virtual environment intended as meeting room. But what might that be?

It is not clear up to now whether a scenario with all three characteristics can be included into a virtual institute in a manner not too artificial. Perhaps, the access to the database could be channeled exclusively through a virtual meeting place (while requests to the database and their results as such need not be in 3D). This scenario reminds us of the foyer of a library: whereas the main function of the library – getting access to the information in books – can be performed quite well (or actually: even better) without physically meeting other users of the library, the foyer offers a place for unplanned encounters, e.g., while waiting for the books or studying some conference posters or ads on the wall. In the reading halls and catalogue rooms, silence is demanded since point (*iii*) above does not hold there (nor does the second criterion). The foyer, in contrast, is usually rather noisy with talk, and the third item is obviously in function. Even criterion (*ii*) seems to be at least partially in reign for the foyer exactly because the work that one intends to do when passing the foyer is actually associated with other rooms and may be postponed for a quick chat on the way.

Finally, we should not forget another aspect highly relevant for informal computer-mediated communication in a 3D environment. Much of the communication happening during "kitchen encounters" is of non-verbal nature. From facial expression to intonation, from body language to eye contact, many expressive background signs enrich the verbal foreground and have an enormous impact on the communication. It is presumably the missing of this additional level of communication that makes scientists hesitate to publish unfinished scripts even in the small circle of colleagues while discussing the very same drafts without any reluctance face to face [SMITH *2002, 59*]. It may, thus, be doubtable whether relatively rigid avatars and written chat are already immersive enough for a meeting place in the sense intended. The integration of *viva voce* communication and even video conferencing must certainly be an option at least.

Let us come back to the distinctions of Table 5. Quite obviously, immersion and communication have very different importance for each of those categories of tasks. Table 6 associates the preferred functionality with the four types of tasks as a kind of résumé of the preceding discussion. Whereas solitary communicative tasks can be best performed without overloading the interfaces with too much "virtuality", solitary immersive tasks and cooperative communicative tasks call for the specific forms of virtuality provided already by pure virtual realities or straight virtual communities (in the wide sense), respectively. Solitary immersion is not too important for our case study so far and adds essentially an aesthetic moment for public relations. Communicative coop-

Table 6	immersive task	communicative task
solitary task	*functionality of virtual realities*	*classical single-user interfaces*
cooperative task	*virtual reality with multi-user communication and other coordination functions*	*functionality of virtual communities (without disguise)*

eration is, in contrast, highly relevant, and all the techniques of virtual communities could be applied (apart from their option of disguise).

Only the class of cooperative immersive tasks, which are essentially associated to informal meetings, demands the full combination of virtual reality with the synchronized interaction of virtual communities. The sign character of the pictures setting up the virtual environment has to be suppressed up to a high degree. In accordance with our "kitchen encounter" metaphor, the setting of this part of a virtual institute is quite critical: we have seen in this section that there are still many questions to be answered concerning a proper and continuous motivation of the members to employ the functionality offered in this respect. For support, a "natural" integration of the other functions is desirable.

5.3.4 Conclusions

The two application examples of immersive systems have demonstrated that the balance between deceptive mode and symbolic mode in the uses of interactive pictures depends stongly on the actual sub-tasks, and hence may vary with the user's intentions. While the tasks in the example of virtual architecture considered above call for controlling the symbolic mode of reception of the pictures used in the immersive system by means of pictorial estrangements and thus allowing for more or less deception by its users, the function of the virtual institute depends in parts on pure deception. Its other functions however cannot be granted by enforcing the symbolic mode of reception, but need a totally different presentation. An immersive system may even hamper those tasks.

The dependencies in the use conditions of immersive systems follow directly from the concept »picture« and, thus, are linked to the underlying data structure »image«: We indeed have to consider more than just the computerized version of the picture vehicle (i.e., the data type »image« only formalizing pictorial syntax). Integrating some of the semantic aspects by a hierarchically structured geometrical model already appears almost trivial. But mirroring in the beholder models pragmatic aspects of the user's intentions plus the effects to be expected by the pictures generated is also a necessary part of the data structure. It forms the core of any more autonomous control of the balance between deceptive and symbolic mode in the reception of computer-generated pictures even in the case of interactive *trompe l'œil* mostly received in deceptive mode.

5.4 Another Border Line Case: Mental Images

Computational visualistics deals essentially with pictures in the usual sense, i.e., with entities with a material carrier that is visible to different persons (at least in principle). A common use of the expression 'image' however does not refer to such entities, and we have to ask whether such an extended sense of images may still fall into the field of

interest of computational visualists: *mental* images. That question is answered positively by means of an exemplary case study employing results of our considerations about the data type »image« in an application dealing with mental images in the context of the pragmatics of objective descriptions of spatial events happening far away.

5.4.1 An Example Task: Understanding Reports From Absent Spatial Events

A typical example from our ordinary life is the task of a radio sports reporter: apart from emotional effects, which we shall ignore in the following, he has to give to his audience a (more or less objective) verbal description of the development of spatiotemporal configurations his audience cannot perceive by themselves. The reporter's behavior is often explained by means of reference semantics: the meaning of the utterances forming his report is understood as being anchored in his (visual) perceptions, as introduced in section 4.3.1.

While examining from the perspective of computational pragmatics the verbal activity of a radio sports reporter who describes objectively what he sees happening on, e.g., a soccer field, our focus of attention is directed essentially to the following three general problems: the speaker should be sure that any assertion of his description can be understood in its particular context by the listeners assumed with respect to reference, plausibility, and adequacy:

Reference: first, a listener should be able to correctly and uniquely identify those objects in the common discourse universe that are used by the speaker to anchor contextually the assertion. Ambiguities in the literal meaning of definite noun phrases and the reference of corresponding pro-forms must be resolvable. For example, the correct use of under-specific definite descriptions, like '*the* penalty area', or '*the* defender', is to be controlled by the speaker's anticipation of the listener's understanding.

Plausibility: even if the listener is able to anchor the utterance correctly in the context, she may fail to understand the assertion since the new information communicated is not plausible for her in the contextual situation. Since the new information essentially transforms or further restricts the context of the assertion in question, such a rejection due to lacking plausibility may occur if the additional restrictions are incompatible with the given context. The speaker has to anticipate whether the listeners are able to integrate the meaning of a continuation of the description presently planned into the understanding assumed so far. On the verbal surface, this shows essentially in what has been recognized but is not said.

Adequacy: finally, under the assumption that the assertion communicated also is plausible for her, the listener may draw implications that the speaker does not want her to draw. In the case of an objective description, the question is whether the listener's conclusions are adequate with respect to the events observed by the speaker. In particular, it is an interesting task for the speaker's anticipation of the listener to initiate – under the general restriction of economy [GRICE *1974*] – additional information only in cases where it is necessary to keep the listener's understanding adequate: on the verbal surface, such additions may be found in grammatically optional, locative expressions like 'she receives the ball *at the left penalty spot*'; here, again, some further consequences appear in what is left out in the description actually produced.

In order to explain how a listener understands the report grounded in the visual perception of the speaker, the listener is usually assumed to have constructed a visual men-

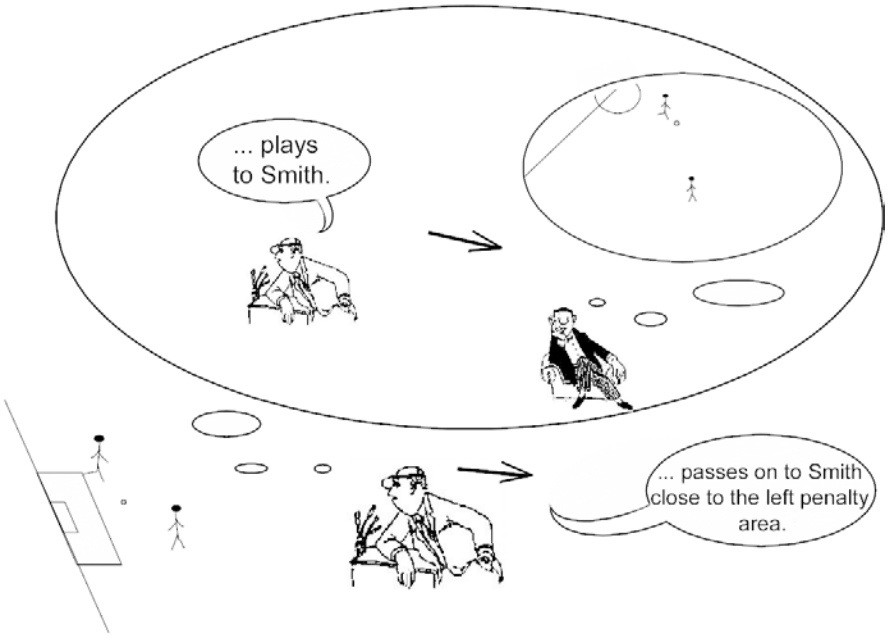


Figure 121: Sketch about Mental Images in Explanations of Reference Semantics

tal model – a "mental image" – which substitutes percepts of scenes not perceptually present. "The radio reporter has solved his task only if he describes the reality of a sports event so vividly and obviously to the listener that the listener believes she sees that reality" a German linguist wrote [DANKERT *1969, 94*]. The essential claim is such that the spatial implications the listener is able to draw from the reporter's descriptions can simply be "seen" in those visual pseudo-percepts: the listener would be able to "see in her mind's eye" that a certain player stands to the *left* of the *opponent* penalty area after being merely told that that player is *beside the* penalty area. That particular understanding is assumed to be the only consistent way for her to continue the contextual mental image. In doing so, she resolves ambiguities both included in the meaning of the preposition used and in the reference of the noun phrase, but without using spatial reasoning in the explicit way described in Table 2 of section 4.3.1.3.

It should be rather clear that such a conception opens a way to solve the problem of integrating the need for referentially anchoring semantics with the idea of partner modeling: first, the mental image would allow the listeners to anchor the speaker's utterances referentially in analogy to the speaker himself. Although mental images are not precisely visual percepts, they are conceived of as being very close relatives that can be used as substitutes. Thus, we could assume the very same kind of semantics to be used both by the speaker and his audience. The listener model of the speaker correspondingly has to deal with mental images, as well. The speaker, then, is thought of as taking into account the mental image his listeners are able to construct in accord to his utterances: if this mental image does not fit to his communicative intentions, he has to change his utterance plan accordingly (Fig. 121). Before presenting the computational example, the function of mental images has to be elaborated a bit further.

5.4.2 On the Cognitive Function of Mental Images

> We do not have to ask what mental representations are, or what happens when we imagine something, but how the expression 'mental representation' is used.
>
> [WITTGENSTEIN *1953, §370*]

The remark of WITTGENSTEIN above holds for 'mental images' (or 'pictorial mental representations') in particular. That is, we should ask: How and under what circumstances do we speak meaningfully while using the expression 'mental image' or one of its synonyms?

Most contemporary cognitive theories agree on that a listener while concentrating her attention on a sportscast on radio usually *imagines* the described spatio-temporal configurations. More precisely: the concept »mental image« appears in a specific sort of explanations of an aspect of what happens "mentally in" the listener of a radio report: it is proposed that, in order to understand the description, the listener has to represent, i.e., to bring to her presence, *in a concrete, "sensible" form* what is described. Since the description is primarily anchored referentially in something seen, it is assumed that the listener imagines the scene in a form that substitutes a corresponding visual perception.

This idea in fact originates from the mentalistic framework of the Philosophy of Enlightenment. In the dawn of this position, R. DESCARTES and especially J. LOCKE understood a concept to be a mental image, or more precisely, a *prolongated perception* of a corresponding particular that serves as a *prototype* for similar particulars. However, this interpretation ran quickly into severe problems [ROS *1990, Vol. II, 55ff.*]. Integrating parts of this idea with G. W. LEIBNIZ's conception of a concept to be a human faculty, i.e., a *mental program* for recognizing corresponding instances, I. KANT in the heydays of the Philosophy of Enlightenment presented an elaborated theory of a twofold mental construction: first, he considers a human faculty of constructing concepts which, second, themselves are mental faculties to construct *intuitions*, i.e., mental representatives of instances, or more colloquially: mental images [KANT *1781, B741f./A713f., A105 & B180/A141*]. KANT's second step, the construction of mental images of instances, was resuscitated in contemporary cognitive science by P. N. JOHNSON-LAIRD under the name of *mental models* [JOHNSON-LAIRD *1983*]: in the mentalistic tradition, the context of an utterance is interpreted as a mental model; the nominations of the utterance under investigation are expected to identify elements of that model; its predication is used to communicate an additional distinction (with respect to a concept). By means of that faculty, the contextual mental model is transformed into (the perception of) a concrete instance of that concept. Thus, all implicatures of the application of the corresponding distinction in the given context have to be present in the resulting mental image. For KANT, those faculties for constructing or revising mental models are autonomously created – "synthesized" – by the human mind, as well. More precisely, he refers to synthesizing a completely new field of concepts by combining several given but originally unconnected fields of concepts. The introduction of the rational numbers as a combination of two (sets of) integers (counter and denominator), which we already met in section 4.3.1.2, can therefore be viewed as a synthesis in the sense of KANT.

The crucial question of the traditional conception of mental images is the *privacy* ascribed to them. The most obvious consequence of the assumed privacy of mental representations is, that there is no way to determine whether or not an instance really is present "in some other mind". We need not share the mentalistic fundament of LOCKE, LEIBNIZ, and KANT: following instead the linguistic turn indicated by the quote of WITTGENSTEIN, we shift the focus of our attention from the construction of a concept

understood as a private mental entity to the explanations we could give for the explanative power of a concept conceived of as an abstraction of verbal behavior: in order to explain why the concepts of a certain field can be used to explain assertions with corresponding predications we could remain *within* that field of concepts, employing merely its meaning postulates. Or we could additionally consider the constituting schema of that field: its internal structure is then viewed as combined from those of other fields. Exactly these two types of argumentations have already been discussed in section 4.3.1.2. Recall here in particular the application of the field-external argumentation concerning the synthesis of the concept of (sortal) spatial objects from the fields of contextual geometric Gestalts and abstract part-whole relations: that argumentation has allowed us to explain visual perception and its role for pictures in section 4.3.2.

We can interpret the meaning postulates with respect to spatial concepts, like the rules of transitivity of the concept 'being in' or the rules of conversity between the concepts of the projective prepositions, simply as expressing the internal structure of that field. We may use them to logically explain the adaptation of the context resulting from a new spatial assertion: we describe the context – i.e., what we assume to be the common knowledge of speaker and hearer – by a set of sentences with spatial predications. Meaning postulates corresponding to the predication of the new utterance are used to add further statements in the syllogism-like manner of spatial reasoning, thus making explicit the implicatures of the utterance in that context. Let us call this procedure the *horizontal* dimension of explaining the understanding of utterances about spatial entities (Fig. 122 upper half): the context is based totally on the analysis of what was said before, and its revision takes place within merely one field of concepts.

We may also view the meaning postulates as parts of a more ambitious argumentation: for example, we may say that the concept 'being in' is in certain cases transitive and in other not, *because* it is introduced in a particular way on concepts of other fields with their characteristic internal specifications. Then, we focus on the two fields of concepts that we conceive as crucial for implementing the field of spatial objects: the field of configurational Gestalt concepts (geometrical level), and the field of functional part/whole concepts (meronomic level; Fig. 122 dotted arrows). Founding the properties of spatial concepts synthetically thus means to explain them with the interaction of the properties of the geometrical and the functional field. Let us call this aspect of explanation the *vertical* dimension, since the synthesis constructs higher, i.e., more complicated fields of concepts, from simpler ones. Any set of propositions – or context – of the spatial field of concepts can be vertically explained as a synthesis of a set of propositions of the geometrical field with a set of propositions of the functional field: each spatial proposition predicating on a sortal object is projected to configurational propositions predicating on the – perceptible – Gestalts of the sortal objects, and functional propositions predicating on its meronomical relatives.

The geometrical level provides the concepts used to describe the (essentially visually) perceptible attributes of sortal objects. As introduced in section 4.3.2.3, the interpretation of a context of the geometrical field as a projection of a corresponding context of the spatial field (with an appropriate meronomical presupposition) can be viewed as an explanation of visual perception: the geometrical field providing the visual aspects of space is the same as the one determining our considerations of pictorial syntax. Although a mental image is not exactly a sign in the same sense as a material picture, the analogy of using the geometric projection of a spatial context motivates us to call it an image, as well. We therefore may apply at least some aspects of the data type »image«

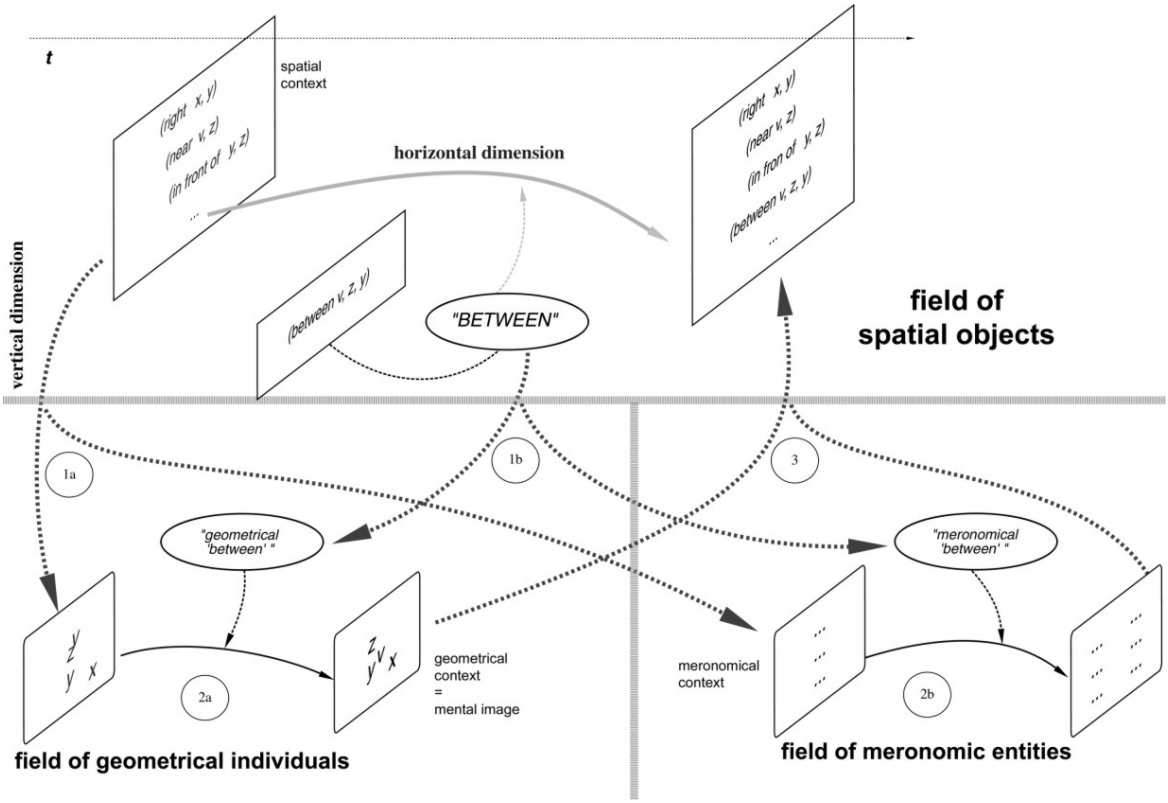


Figure 122: Horizontal and Vertical Dimensions of Explaining the Understanding of Spatial Assertions

when dealing with mental images computationally, in particular its syntactic and semantic aspects.

With this, we are finally able to present a more clearly elaborated version of understanding spatial reports in the framework of reference semantics: the revision of the spatial context by means of the predication's concept – e.g., 'to be in' – is partitioned into three steps (Fig. 122): first, the proposition of the utterance (including the context) is transformed by following the schema of the spatial field into a corresponding structure of sets of propositions of the lower fields (1a & b). Second, the revision of the context by means of the spatial concept of the predication takes place on the lower fields (2a & b): coordinated by the schema of sortal object constitution, the corresponding projections of the spatial context are revised by those concepts of the lower fields implementing the spatial concept in question. Third, the resulting partial understandings – especially the derived context of the geometrical field, called 'a mental image' – are synthesized back to form the spatial context for the subsequent utterance (3); the resulting context includes the spatial implicatures of the utterance in question. This step is equivalent to the goal-driven phase of perception and may be directed by pragmatically motivated focusing strategies. With this schema, a corresponding computer model can be designed.

5.4.3 Building a Computer Model

A corresponding integration of the vertical and horizontal dimensions of explaining spatial cognition is exemplified by the system SOCCER of the project VITRA [ANDRÉ *ET AL. 1988*]: in this case, the exemplary radio sports reporter from the beginning is considered. The explanation of the visual perception, which is part of the foundation by

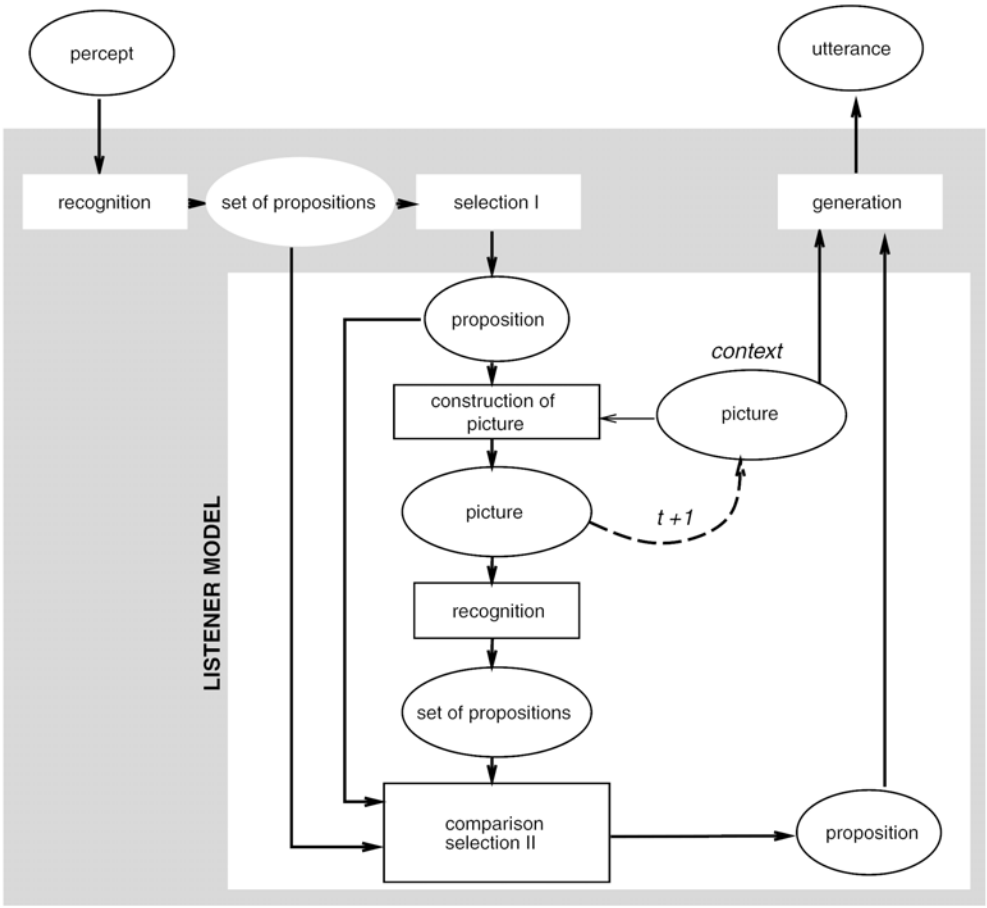


Figure 123: Architecture of the System SOCCER with its Listener Model ANTLIMA

reference semantics of the utterances of the radio reporter, follows the exemplary line given in section 4.3.2 up to the field of spatial concepts. Simplified versions of the concepts underlying static spatial relations, like »being in«, »- at«, »- near«, »- to the left« holding between a reduced version of sortal objects are determined. The concepts of spatial events, like »doing a double pass with«, are additionally defined as a temporal sequence of phases during which certain spatial relations hold. From the resulting sets of spatial propositions, some are finally chosen to be communicated and transformed into a corresponding verbal manifestation (Fig. 123):

S1: Miller, the defender, stands just left to the penalty spot.

S2: Miller gets the ball and runs with it close to the centre circle.

As was mentioned in sections 3.5 and 4.4.2, any adequate theory of communication explaining the behavior of a speaker also has to consider the audience in a particular way: the speaker has to be conceived of as somebody who also sets himself in the position of his audience. He has to play anticipatorily its role in the language game in order to really communicate. In VITRA, this demand is answered by means of the listener model ANTLIMA: we focus here only on the static spatial relations, as in sentence S1, although spatial events as in S2 are dealt with accordingly, as well. The understanding of the audience is modeled with the three steps described above:

First, the proposition of the (planned) utterance is projected to the lower levels implementing the spatial field: i.e., restrictions of the spatial interaction with other objects

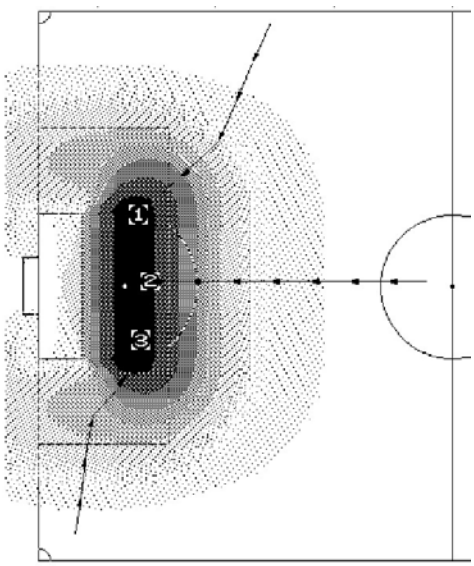


Figure 124: Visualization of the TyPoF for a player being in front of a penalty area, and approximation paths for several contexts

are transferred mainly into restrictions of the locations of the objects (plus the part-whole aspects of the objects involved); this transformation – the schema of the corresponding spatial concept – is encoded in ANTLIMA by means of functions called 'TyPoFs',[98] which are already applied for recognizing spatial relations: they can easily be viewed as the characteristic functions of the fuzzy sets of situations to be described by the corresponding relation (cf. again Fig's 66 & 67, Sect. 4.3.2.3).

Second, the context of the planned utterance is revised on the lower level, i.e., as a mental image: the locations of the objects are chosen by means of a hill-climbing algorithm ruled by the TyPoFs and depending on the contextual positions. Figure 124 illustrates the influence of three different geometrical contexts (starting positions) on the location selected, namely 'to be in front of the penalty area'. The hill-climbing algorithm determines maximally typical positions for all objects localized with respect to the geometric restrictions given by the predication. Therefore, the image construction concretizes the consequences of an additional proposition to the given contextual image – an implicit type of spatial reasoning. If an image can be constructed with highly typical positions for all restrictions, the utterance under consideration must be rated plausible in the given context.[99]

Third, the schemata of the spatial concepts (object models, TyPoFs, and definitions of spatial events) are applied to (re)construct the context on the level of the spatial field: this finally renders explicit the implicatures included. Another set of spatial propositions is the result.

That set modeling the anticipated understanding of the audience has to be compared in the listener model with the understanding intended by the speaker, i.e., what has been actually perceived: the differing propositions are used in an anticipation feedback loop for an enhancement of the propositions to be effectively uttered (cf. again Fig. 123, and Fig. 84, Sect. 4.4.2.2).

Note, that the image constructed – i.e., the image the speaker anticipates the listeners can construct when told the proposition in question – cannot directly be compared to the set of propositions describing what the speaker has observed. A first guess might be to use the percept instead – after all, the audience should have a mental image corresponding to the speaker's percept. Percept and mental image are assumed to be of the same type, so that the comparison can be done syntactically. Unfortunately, such a solution is not exactly plausible. That conception does not take into account that the speaker's communicative intentions are – even in the case of an objective description – not identi-

[98] 'TyPof' is a speaking acronym for 'Typicality Potential Field', alluding to its use in a gradient search: it *tips off* the maximally typical positions falling under the spatial concept in question.

[99] The resulting image is later used as the starting point for constructing the image for the utterances planned next, and also to check whether a noun phrase to be employed in that utterance denotes uniquely an object in that imaginary visual field of the listener solving the question of reference.

cal to the speaker's "raw percept": it is the set of (spatial) propositions reflecting what the speaker has recognized in the percept, which has to be considered. Even if we assume that the comparison between the percept and the mental image could be used – for example by means of the distance between the two incarnations of an object in the two images – we still have a serious difficulty: are all differences really equivalent? Imagine a soccer field with two balls – a black one representing the position perceived by the speaker, and a white one representing the position anticipated by the image construction. Let us assume furthermore that the two balls are in one case about one foot apart somewhere in the middle of the field away from any landmark, and in another case – with the very same distance between each other – on different sides of the outside line (Fig. 125). It should be obvious that in the first case, the difference is not considered essential, and correspondingly should not trigger a reaction in the listener model. However in the second case, the two positions are different: if the white ball is the one outside the field, the listener model has in fact predicted that the audience falsely understands that the ball is outside of the game: a correction is then highly recommended.[100]

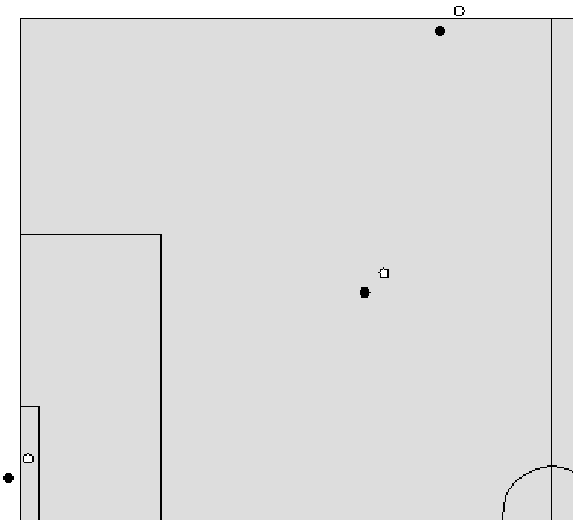


Figure 125: Same Difference – Different Relevance

The recognition component of the speaker model classifies exactly percepts with *essential* differences; it generates the same sets of propositions if two percepts do not differ essentially. Therefore in the listener model with mental images, the very same "cognitive abilities" are employed with respect to the mental image in order to generate a propositional description of what the audience (at least presumably) is able to recognize in its mental image. That set of propositions can easily be compared to the analogous set of the speaker providing the means for dealing with the problem of adequacy mentioned above. Thus, the sequence of recognition and secondary selection based on the anticipated mental image reflects exactly the speaker's own activities with respect to his percept: recognition and primary selection. The analogy of the »seeing by one's mind's eye« and the »seeing by the physical eyes« becomes even more plausible: as was said before, it is believed that the listeners can "see" the consequences of integrating a new proposition in the contextual knowledge in the mental image.

As is demonstrated in Figure 126, the spatial restrictions holding for an object simultaneously (e.g., during an event phase) can be easily combined on the level of TyPoF's. Only if the combination is consistent, the resulting typicality field has a maximum close to the ultimate value. Furthermore, the context-sensitivity of the algorithm for finding the maximum of the typicality distribution as demonstrated in Figure 123 adds another advantage when considering spatial events: the positions of the objects at consecutive

[100] See also again Figure 85 (Sect. 4.4.2.2): the argument used here is also valid for the comparison step in the listener's anticipation feedback loop: the pictures can only be compared with respect to a particular "reading", not as such.

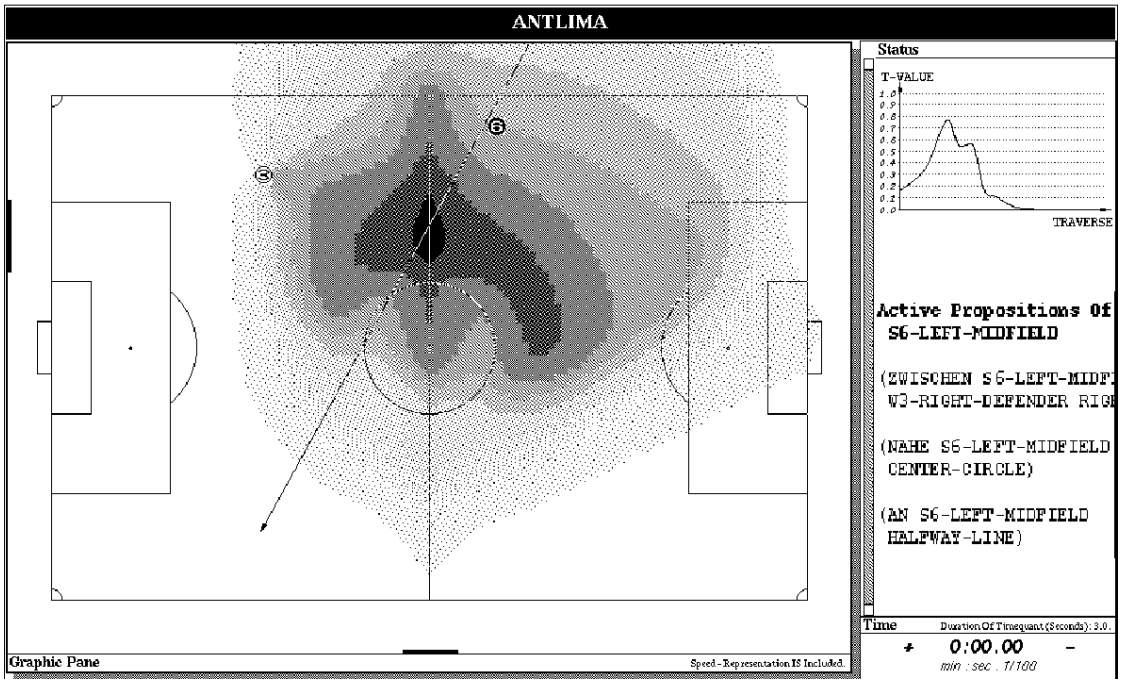


Figure 126: Consistent Combination of Spatial Restrictions: the TyPoF for "being simultaneously between the player No 3 of the White Team and the right goal area, near the center circle, and at the halfway line" (upper right: cross section through the typicality field along the arrow in the main panel)

moments of an event phase are developed in a cinematographic procedure taking the position at *t-1* as starting position for the gradient search for the position at *t*.[101]

The further uses of the difference propositions in the listener model are not directly related to the data type »image«. They have been described elsewhere in detail [SCHIRRA *1997*].

5.4.4 Conclusions: The Data Type »Image« and Explaining Mental Images

Indeed, the data structure used in ANTLIMA for modeling mental images is a large subset of the data structure »image«: syntactically, the very same restrictions apply. Similarly, the semantic level is identically resting on the implementation relation between sortal objects (as image content and reference) and the geometric Gestalts (of the picture syntax). Only the pragmatic aspects of »image« do seemingly not play an important role in this use of the data structure, as the image content seems not to be embedded in a twofold beholder model in the same way as described in section 4.3. However, the comparison with the "raw percept" – computationally handeled as an instance of »image« – by means of the corresponding set of spatial propositions is structurally equivalent to the relation of the two beholder models. Indeed, the instances of these mental images in the computer can easily be viewed on a screen, too, and are then used as regular pictures by the beholders. Despite the educated opinion that mental images are certainly not pictures, it seems that essentially the same data structure is able to cover both phenomena.

[101] A more detailed description of some other problems associated with concretizing spatial events is to be found in [SCHIRRA *1994, Chap. 10*]. That the method employed in ANTLIMA for positioning objects can also be adapted easily to control the camera positions in a virtual 3D environment, e.g., a computer game by means of verbal orders has been demonstrated by a recent diploma thesis [BERNHARDT *2003*].

This mysterious contradiction is solved quite easily: recall what we have found out about the relation between fields of concepts and data structures: data structures correspond to fields of concepts; and concepts as we understand them are everything that is structurally common to all the explanations of the corresponding predicative expressions. The *concepts* of mental phenomena, including mental images, therefore are not strange entities (although the phenomenon they cover may be strange if the concept is ill-formed): by mentioning such a concept, we simply refer to what is common to all explanations given about the situations a corresponding expression (e.g., 'having a mental image of something' – or rather 'imagining something visually') is used, something that is very well observable by any beholder. Being able to use the data structure »image« in a system about mental images therefore only means that the explanations for using the expression 'picture' are to a significant degree structurally equivalent with the explanations given for using the expression 'mental image' – and that is not strange at all. "The radio reporter has solved his task only if he describes the reality of a sports event so vividly and obviously to the listener that the listener believes he sees that reality": we have met this quote from a linguist about mental images in sports reports above. "In an immersive system, the beholder is forced to believe he sees the reality instead of using a sign" we may add with respect to pictures. 'Believing to be seeing a situational context that is only mediated by means of signs' states the common structure underlying both explanations. Thus, the synthetical relation between image syntax and sortal objects is explicitly focused when bringing mental images into the game of explaining spatial assertions, a game that can also be played without that complications by means of field-internal spatial reasoning alone – "propositionally" . But then we would give up the option to take the reference semantics perspective.

* * *

Four case studies are of course a rather limited set for demonstrating the value of the preceding investigations. Nevertheless, they cover a wide range of applications even up to the strange phenomenon called 'mental images'. And they indicate that developing new programs is only one part of the tasks a computational visualist has to deal with. The other one is: to carefully listen to those knowing the field of application – mostly image scientist of various disciplines – and to reasonably translate their concepts into the formal structures computer scientists are specialists for.

6 Conclusions – Perspectives

Investigating the variations of the generic data structure with the type »image« and its application conditions has been the main goal of the previous chapters. Only a selection of aspects has been discussed – there are too many to be exhaustively included, at least for the time being, as a unified theoretical view on pictures only starts to form. However, we have been able to sketch a generic data structure with a non-trivial set of main ingredients necessary for the foundation of computational visualistics as a unique scientific domain (Fig. 127). Like all generic data structures, many components are only indicated as necessary but are not elaborated. They have to be filled differently for the various sub-types covering one of the multitude of application contexts of computerized pictures.

6.1 The Components of »Image« as Basis of Computational Visualistics

Each picture has a vehicle that determines the syntax: in the generic data structure, a corresponding type is contained consisting of a geometric base structure and colors or textures as marker values. The concrete realization of this component – e.g., as a pixel matrix with fixed resolution (sufficient for many tasks of image processing), or as a generative sub structure with an explicit zooming operation – depends on the task at hand.

Pictures also have meaning aspects in general, though this aspect is already much more complicated than syntax. In the generic data structure, the central semantic component is a data structure containing the formal equivalent of the object constitution relation between the concepts for sortal objects (as the primary target of an image's meaning: the »picture content«) and their geometric and meronomic projections. The representation relation Rep and its inverse relation have been discussed extensively in section 4.3. They essentially relate syntactic and semantic components. In many cases, only the geometric projection of the sortals is actually considered (e.g., as the 3D models used in computer graphics). Although the sortal concepts form the semantic core of pictorial semantics, associated individual sortal objects are important, too: they are particularly needed if an identification of the depicted object with a sortal from another context is to be considered (»picture reference«).

Derived forms of meaning relevant especially for information visualization arise by means of an additional metaphorical projection from the sortal concepts to a target domain. This path of derivation can be followed from the simplest form (non-visual properties of sortal objects are visualized: projection from sortals to sortals) through intermediate levels (projection from a completely different field to sortals; projection from an arbitrary field to the instantaneous geometric projection of sortals) to the extreme (direct projection to the syntactic aspects, i.e., pure 2D geometry) with its special form highly relevant in the artistic domain (the identity projection resulting in no real semantics).

As we have seen in section 4.4, pictorial pragmatics is an even more complicated field. In a first step, image uses can be associated with intentions and goals of the picture users. They are related to the picture content (sortal or target domain). Those intentions and goals are usually arranged in something like sign act games and determined by rhetoric relations. This is the primary pragmatic component of the data structure.

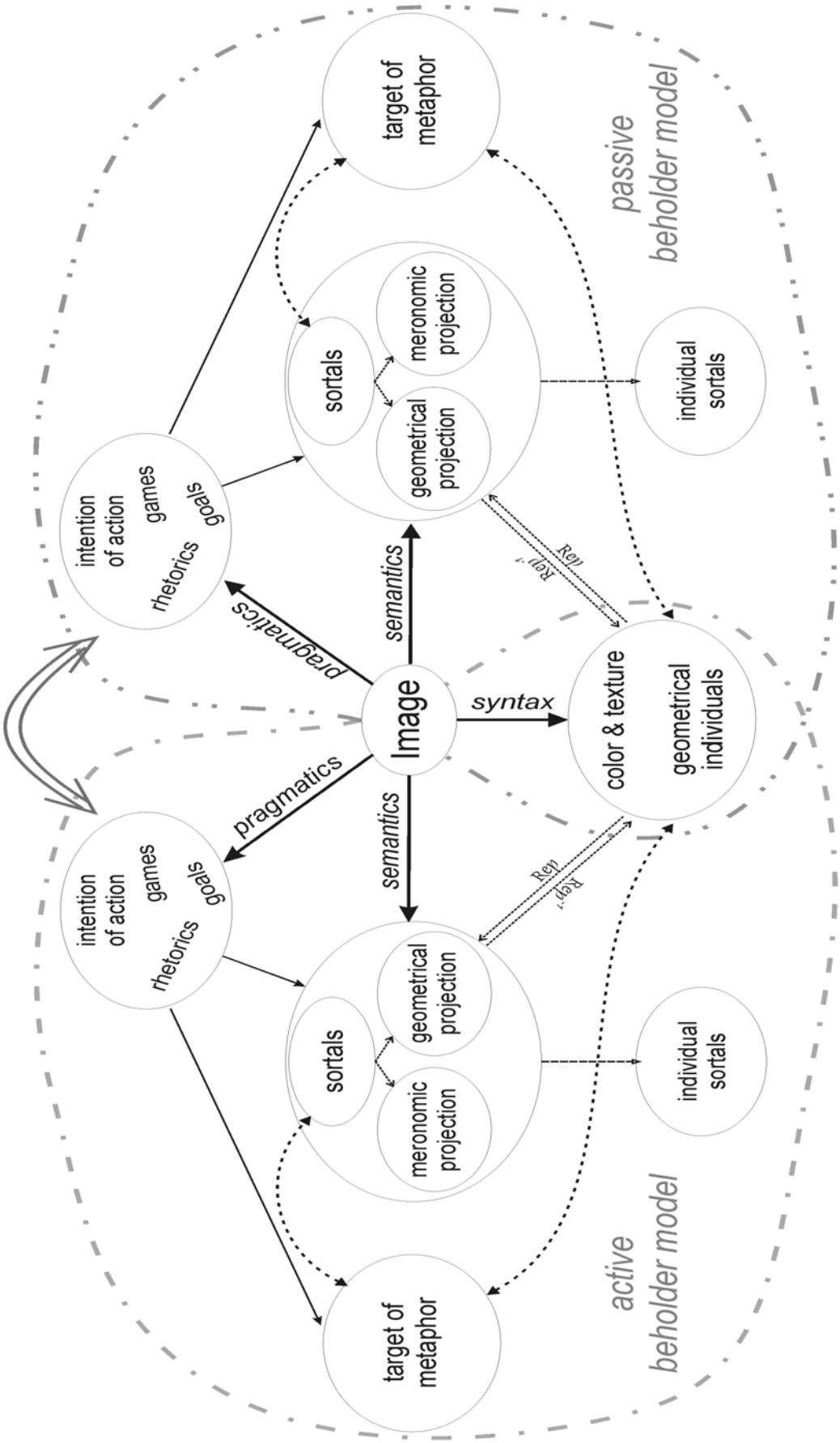


Figure 127: The Generic Data Structure with the Type »IMAGE«

However a refined analyses has lead us to the conclusion that in fact the semantic and primary pragmatic components have to be duplicated by means of the active and passive beholder models: only in the mutual relations between the intentions of the sender and the receiver – and also between the picture contents for the former and the later – pictorial pragmatics can be adapted to the needs of interactive systems with tele-rendered images.

Merely the picture vehicle is given simultaneously to both participants in a pictorial sign act – indicated in Figure 127 by means of the overlapping regions of the two beholder models. That a picture is so often mixed up conceptually with its vehicle is perhaps motivated by this fact. However, it is only the integration of the two perspectives on semantic and pragmatic aspects that actually constitutes – even in a soliloquial use – the picture *per se*, i.e., the data type in the center of Figure 127.

Figure 127 can be read as a coarse map of the realms of computational visualistics, a structural context builder summarizing what the whole of this text has tried to elaborate in a more extensive though linear manner. Scientists in this domain may take it as an orientation aid; their actual task is to work out more precise maps of some of the regions sketched here, to find passageways and short cuts to be used under certain conditions.

Let us keep in mind the methodological roots of computer science being simultaneously a structural science and an engineering science: the structural perspective has certainly dominated our discussion. This is of course partially due to the fact that technical implementations can hardly be incorporated adequately in written investigations. But this by no means intends to indicate that the engineering aspect is not relevant here: important work in computational visualistics is dedicated to technical implementations of the data structure »image« in the aspects helpful for the task at hand. However, those implementations always rely on a properly structured background: the implementation relation between data structures is primarily an abstract constitution relation, and only secondarily transferred to technical artifacts.

6.2 Computational Visualistics in Education

In section 1.4, the conception of a “new engineer” has been sketched, a conception that also governs the self-image of computational visualists. In that perspective, engineers are successful only to the degree to which they are able to adapt their work to the needs of clients, to incorporate the expertise of those clients, and to communicate their solution to the clients. In this sense, the extended chapter 3 about image science is as essential a part of the foundation of computational visualistics as is the analysis of the components of the data structure »image«.

The future of computational visualistics as a new discipline in computer science over and above the well-established fields of computer graphics, information visualization, image processing, and computer vision depends essentially on the success the discipline has with respect to the two roots of the Western academic system: research and education. While chapters 4 and 5 have dealt exclusively with aspects of research, computational visualistics can also be established in education successfully and in a unique manner that differentiates it from other degree programmes in computer science, and gives rise to a special identity of those having finished the programme: the “image” of being a computational visualist.

6.2.1 An Example: Structure of Computational Visualistics in Magdeburg

In the fall of 1996, the degree programme 'computational visualistics' has been introduced by the Department of Computer Science of the Otto-von-Guericke University at Magdeburg [SCHIRRA *1999*]. Its backbone is formed by three columns (cf. again Fig. 41, section 3.5.4):

- *Method*: the algorithmic handling of pictorial data structures,
- *Reflection*: aspects of dealing with pictures in the humanities, and
- *Application*: an example domain of practical visualistics beside computer science.

The main column covers computer science. Students learn about digital methods and electronic tools for solving picture-related problems: they dedicate about 60% of their time to study the variations and application conditions of the data type »image«.[102] This technical core is complemented by courses in general visualistics in theory and in practice: about 25% of their credit hours, students are occupied with theoretical aspects of pictures discussed in the humanities. Besides learning about the "traditional", i.e. not computerized, contexts of using pictures, and the theories developed in disciplines like psychology, educational and political sciences, philosophy, and also industrial design, students here intensively practice their communicative skills. Furthermore, an application subject, which covers about 15% of the credit hours required, gives students an early opportunity to apply their knowledge: they learn the skills needed for cooperating with (e.g., carefully listening to) clients and experts of an example customer field such as medicine.[103]

The tripartite division follows current analyses in educational science [GIRMES *1997*]: the competence we want to teach is not the result of merely acquiring certain techniques and methods, i.e., a certain technical *repertoire* our students are provided with after graduating. That repertoire is always used within a particular *context of application* with its proper peculiarities. Moreover, there are usually experts in that field of application who have their own traditions and methodologies. They act as clients of computational visualists, and have little interest in being patronized in their own field of expertise. *They* specify the problems to be solved by computational visualists, as well as the criteria of quality for the solutions proposed. A third factor of competence is defined as the ability to *reflect the conditions* of successfully applying a method of the repertoire in the given context, e.g., a visualizing technique in material science: although a method might in fact be successfully applicable to reach the goal, there can be conditions in the context of action that argue against the application of that particular method (negative side effects).

6.2.2 Mental Imagery as a Preview Criterion for Study Success

In a time with rather limited academic ressources, restricting the access to a programme to those with a high probability of being able to successfully finish it is a legitimate thought. This is even more true for a new degree programme since potential students

[102] In order to ensure accreditation with the German professional bodies of computer science, at least 50% of the course work has to be within the core of computer science.

[103] The students choose one application domain at the beginning. Presently there are four options: medicine, material science, construction and production, and electrical engineering (pictorial information systems).

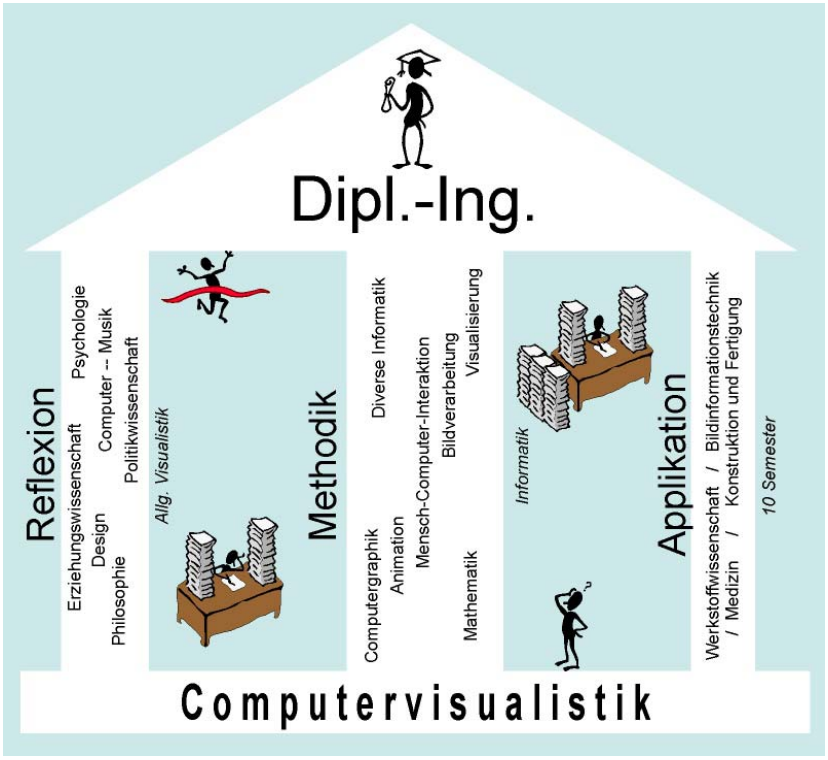


Figure 41 – Reprise: "The Three Columns of Computational Visualistics"

may have strange ideas about what to expect. An elaborate preview for their study success helps them to understand their strengths relative to the programme's intentions. It also helps the university to focus its efforts and to diminish the "loss" of students (almost 40% in computer science programmes in Germany; cf. [HEUBLEIN *ET AL. 2002*]).

For computational visualistics, the ability to mentally imagine may be one of the key criteria for such a preview. Psychologists have invented an extended set of tests for different aspects of mental imagery, for example, the ability to anticipate mentally object rotations. The general relation between mental imagery and computer use is also well documented (e.g., [DE LISI & CAMMARANO *1996*]). There is, however, a severe drawback: it has been repeatedly reported that women are less successful in tests on mental imagery than men (e.g., [MAIER *1996*] or [QUAISER-POHL *2001*]). As a kind of "gender effect", the different conditions of socialization and the different options for experiences for boys and girls may be the cause [NEWCOMBE & DUBAS *1992*].

Women are massively under-represented in computer science: in average, only about 7 to 8% of the new students are female in Germany [SCHINZEL *ET AL. 1999*]. The situation is similar in the other fields of engineering: less than 6% women start a degree programme in electrical engineering, for example. Nevertheless, for more than 6 years in sequence since the programme has been installed, about a quarter of the beginners in the computational visualistics programme at Magdeburg were female (some years even more than 30%).

We obviously cannot have an interest in installing a preview test that handicaps women. It is however important to understand to what degree mental imagery is still a useful criterion despite the usual gender-specific differences. From the psychological perspective, there is also need for further investigating the relations between using computers, gender, and mental imagery causing those differences. Therefore, two psychologists have investigated on my initiative how female students of the computational visu-

alistics programme at Magdeburg differ with respect to mental imagery from their male colleagues, from students in other degree programmes of computer science, and from those in other disciplines in general. To start with the major result: computer visualists came out most interestingly as one of the very few groups reported in literature on mental rotation that have no (or extremely small) differences between women and men [QUAISER-POHL *ET AL. 2001*].

6.2.3 An Empirical Investigation

Velocity and precission of mentally rotating three-dimensional objects (presented by means of pictures) has been recorded with the standardized Mental Rotation Test (short: MRT [VANDENBERG & KUSE 1978]): each task consists of five pictures of objects constructed of simple blocks, the first of which appears again in a rotated version in one of the other four pictures (rotation in the picture plane). 24 tasks have been presented to the subjects in two test periods of three minutes duration each, separated by a break of two minutes. For the ability to mentally visualize, a test called "Cuts" (German: "Schnitte" [FAY & QUAISER-POHL 1999]) was employed. In this rather compicated test, the subjects have to imagine simple three-dimensonal geometrical entities (surface only) being cut by a plane, and report the resulting two-dimensional cut figures possible. The test consists of 17 tasks. Furthermore, specific experiences with and attitudes to computers prior to their study have been asked. The influence of other learning experiences has been taken into account by means of the courses selected in school and the relative success in their degree programme so far.

In general, computational visualists have significantly better results than the other students.[104] That is also the case for the women only, in particular with respect to MRT.[105] Descriptively, there was an advantage of male students for all disciplines considered; however, this difference was *statistically not significant* for computational visualistics – in contrast to the other degree programmes of computer science considered.

In another part of the study, older students of computational visualistics have been confronted with a more complicated version of MRT (rotations in the picture plane *and* around the vertical axis: MRT-C), and the test "Cuts": the results are similar. In particular a positive correlation has been found between the number of semesters and the performance in the tests indicating that mental imagery and especially the ability to mentally visualize have been practiced during the study.[106]

The experiences with mental imagery a person has by means of spatial activities are a good option to prove the effects of practicing: therefore, the subjects of our tests have been asked how often they have participated in certain leisure activities that rely on spatial imagery.[107] We have analyzed the relation between the amount of prior experiences and the results on MRT by correlating the two features in the different disciplines for each gender separately. The degree of correlation came out as highly depending on gen-

[104] In average, they reached 14.2 points (of 41); that's 3.7 points more than students of psychology, 2.8 points more than other students of the humanities, and 1.2 points more than student of sports.

[105] With 13.25 points, female computational visualists reached significantly better average results than female students of sports (11.55), psychology (9.5) or other humanities (9.22).

[106] While their male colleagues only got 8.53 points, female computational visualists reached an average of 11.7 points (of 17 points) – the highest result reported so far, cf. [FAY & QUAISER-POHL *1999, 25ff*].

[107] This includes "technical activities" (e.g., to repair a bicycle), "artistic activities and needle work" (e.g., drawing or knitting), and "sporting activities" (e.g., tennis).

der: female students that are technically oriented and have more experiences with computers also have better results in MRT. There is no such clear effect for men.[108]

The results of this study indicate that there indeed is a close relation between the ability to mentally imagine, and working in computer science. Furthermore, the different degree programmes in computer science show significant variations with respect to mental imagery in the general level of accomplishment as well as in the gender differences. The alterations in the content of the programmes may be the cause, but also the "image" of the disciplines in the context of social gender stereotypes. Only in computational visualistics among the disciplines investigated, "typically male" and "typically female" disciplines are mixed under a common theme: computer science *and* the humanities, both with respect to pictures. In this combination, the least amount of gender-differences of mental imagery has been found. More empirical studies have to show in which way the choice of the degree programme in computer science and the success in that programme may be determined by relevant talents and abilities like mental imagery. Which specific prior experiences at the beginning of the university education, and what influences from prior learning and practicing are important? And which role do differences play between women and men in the conception the students have of their own abilities?

Female beginners in computer science degree programmes certainly profit of highly developed mental imagery. However, it appears to be also very relevant if they subjectively estimate their own capacities in a positive manner. Women have to ignore the general social stereotype still effective to some degree: that they "are not suited" for engineering and mathematics. In this context, an important task for further investigation remains: does the particular interdisciplinary conception of computational visualistics with its significant claim on social and communicative competences – i.e., abilities usually viewed as "typical female" – help women to start their academic education with a more positive attitude with respect to their own abilities? It is perhaps this indirect path of influence that leads to the unusually good results of that group in mental imagery.

It is clear from these descriptions that a satisfying preview test is still a long way off. Such a preview test needs not be directly linked to restricting access to the degree programme. Indeed, offering an automatized online "self test" version may be a helpful first step. Although the test conditions cannot be controlled as well as usual, such a test allows potential students to get a coarse approximation of their success probability in the programme. Furthermore, the data gained by such self tests may be related to later investigations and the actual study success and duration in a longitudinal study and lead to successive better versions of the previews.

6.3 The Future of an Institutional Computational Visualistics

We here reach the end of our investigation. Let us finish with a few thoughts about the future of computational visualistics as an institution, not a loose collection of image-related investigations in computer science. Although the international perspective is crucial here, only a glimps at the situation in Germany can be ventured at this place.

[108] 112 women and 71 men participated in the study: technical activities correlated clearly with the MRT results ($r = 0.375$, $p = 0.01$) while sporting activities showed a less clear effect (0.253, 0.10). Prior experiences with computers had a marked positive effect (0.251, 0.05) but again, only for women. A negative correlation was significant for prior experiences with artistic activities (-0.248, 0.001).

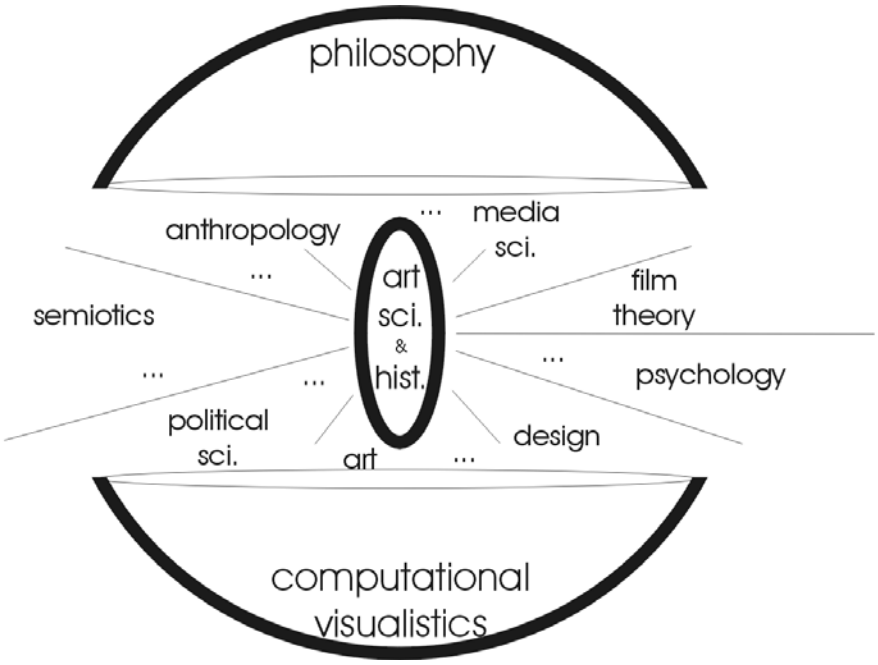


Figure 128: General Visualistics and Computational Visualistics

A comparable step is currently being undertaken for general visualistics in Germany: the Virtual Institute for Image Science (VIB) has already been mentioned above. Another congregation has formed, as well – the "Centre for Interdisciplinary Visualistics" (ZIB): a small number of scientists, each representing a different discipline involved in general visualistics, tries to focus the activities toward an institutional "Bildwissenschaft" in Germany. As a first result, an e-journal on general visualistics has been established.[109]

Computational visualistics is one of the disciplines selected as important for general visualistics.[110] Together with philosophy and art science, it has been assigned a prominent role (Fig. 128). Like an umbrella, philosophy spans the whole range "from above", metaphorically speaking, since it is the discipline that considers argumentations and the proper forms of concepts in general. Art science with its long tradition in art history and its reflections of pictorial sign uses rules the center of visualistics, around which most of the other disciplines are arranged with their external (i.e., not picture-related) parts radiating off. Computational visualistics is a discipline that basically formalizes the concepts »image« developed (following essentially criteria provided by philosophy) in the other disciplines, and thus provides artifacts for "crystallized" (or, for those with a Marxist background: "congealed") argumentations. It spans the other disciplines opposite to philosophy: "from below", so to speak, like a metaphorical funnel collecting their argumentations and transforming them into generally available algorithms. With philosophy as the head, art science as the heart, computational visualistics as the belly, and the other disciplines as eyes and hands, the organism of general visualistics is about to come alive.

[109] IMAGE – An Interdisciplinary Journal on Image Science: www.image-online.info

[110] In fact, the author is one of the founding members of the ZIB and represents the disciplines computational visualistics and cognitive science.

An institutionalization of computational visualistics directly depends on the efforts taken in general visualistics: computational visualistics forms exactly the overlapping region between computer science and image science. In my eyes, it is less marked by external features – a special institute at a university, a journal, a conference series, etc. – although all those means certainly help: the core of an institutional computational visualistics is established in the heads of the people involved. Only if the idea is commonly established that computer scientists in computer graphics, computer vision, image processing etc. work with the same methods at the same subjects, and that in doing so, they essentially take up the arguments of other image scientists to formalize them for others, computational visualistics can act as an autonomous concept with the power to structure the future developments.

With this book I hope to advance that idea.

Appendices

A References

[André 2000] André, Elisabeth: The Generation of Multimedia Documents. In: Dale, R. & Moisl, H. & Somers, H.: *Handbook of Natural Language Processing.* Monticello, NY: Marcel Dekker, 2000, 305–327.

[André 1995] André, Elisabeth: *Ein planbasierte Ansatz zur Generierung multimedialer Präsentationen*. DISKI 108. St. Augustin: Infix, 1995.

[André & Rist 1993] André, Elisabeth & Rist, Th.: The Design of Illustrated Documents as a Planning Task. In: Maybury, M. (ed.): *Intelligent Multimedia Interfaces.* Menlo Park, CA: AAAI Press/MIT Press, 1993, 94–116.

[André *et al.* 1988] André, E. & Herzog, G. & Rist, Th.: On the Simultaneous Interpretation of Real World Image Sequences and their Natural Language Description: The System SOCCER. In: *Proc. of the 8th ECAI*, Munich, 1988, 449–454.

[Asher & Vieu 1995] Asher, Nicholas & Vieu, L.: Toward a Geometry of Common Sense: A Semantics and a Complete Axiomatization of Mereotopology. in: *IJCAI-95*, 846–852.

[Aurnague & Vieu 1993] Aurnague, Michel & Vieu, L.: A Three-Level Approach to the Semantics of Space. In Zelinsky-Wibbelt, C. (ed.): *The Semantics of Prepositions – From Mental Processing to Natural Language Processing*. Berlin: Mouton de Gruyter, 1983, 393–439.

[Austin 1962] Austin, John L.: *How to Do Things with Words*, Oxford: Univ. Press, 1962.

[Barwise & Perry 1984] Barwise, Jon & Perry, J.: *Situations and Attitudes*. Cambridge: MIT Press, 1984.

[Behrens 1901] Behrens, Peter: *Haus Peter Behrens*, Darmstadt: Hohmann, 1901.

[Behrens & Breysig 1901/02] Behrens, Peter & Breysig, K.: Das Haus Peter Behrens. Mit einem Versuch über Kunst und Leben von Kurt Breysig. *Deutsche Kunst und Dekoration* 9:131–194, 1901/02.

[Belting 2001] Belting, Hans: *Bild-Anthropologie – Entwürfe für eine Bildwissenschaft*. München: Fink, 2001.

[Bernhardt 2003] Bernhardt, A.: *Integration einer Sprachsteuerung für die Kamera einer virtuellen Umgebung*. Diplomarbeit. Magdeburg: FIN, Otto-von-Guericke Universität, 2003.

[Bischof 1985] Bischof, Norbert: *Das Rätsel Ödipus*. München: Piper, 1985.

[Boehm 1994] Boehm, Gottfried: Die Wiederkehr der Bilder. In: Boehm, G. (ed.).: *Was ist ein Bild?*. München: Fink, ^{2}1994, 11–38.

[Borgo & Masolo 2001] Borgo, Stefano & Masolo, Cl.: Mereogeometries: a Semantic Comparison. Technical Report 02/01, LADESB-CNR, Padova, Italy, 2001.

[Brachman & Schmolze 1985] Brachman, R. J. & Schmolze, J. G.: An Overview of the KL-ONE Knowledge Representation System. *Cognitive Science* 9(2):171–216, 1985.

[Brooks 1996] Brooks, F.P. Jr.: The Computer Scientist as a Toolsmith (II). *CACM* 39(3):61–68, 1996.

[Buchholz 2000] Buchholz, Kai: Peter Behrens : les fondements théoriques de l'Art nouveau. In: Loyer, François (ed.): *L'Ecole de Nancy et les arts décoratifs en Europe*, Metz: Ed. Serpenoise, 250–261, 2000.

[Buchholz *et al.* 2001] Buchholz, Kai & Latocha, R, & Peckmann, H. & Wolbert, K. (eds.): *Die Lebensreform. Entwürfe zur Neugestaltung von Leben und Kunst um 1900*. 2 Vol., Darmstadt: Häusser, 2001.

[Bühler 1933] Bühler, Karl: Die Axiomtik der Sprachwissenschaften. *Kant-Studien* 38:19–90, 1933.

[Cantor 1874] Cantor, Georg: Über eine Eigenschaft des Inbegriffs aller reellen algebraischen Zahlen. In: ZERMELO, E. (ed.): *Gesammelte Abhandlungen mathematischen und philosophischen Inhalts*. Berlin, 116–117, 1932.

[Chernoff 1973] Chernoff, H.: The use of faces to represent points in k-dimensional space graphically. *J. Am. Statistical Assoc.* 68(342):361-368.

[Clarke 1981] Clarke, B.: A Calculus of Individuals Based on "Connection". *Notre Dame Journal of Formal Logic* 22(3):204–218, 1981.

[Clarke FC] Clarke, D. S.: *Sign Levels – Language and Its Evolutionary Antecedents*. forthcoming: http://www.siu.edu/~philos/facstaff/Clarke/SignLevels/index2.html.

[Cohen 1978] Cohen, P. R.: On Knowing What to Say: Planning Speech Acts. Technical Report TR 118, PhD Thesis, Department of Computer Science, University of Toronto, Ontario, 1978.

[Cohen 1988] Cohen, Harold: How to Draw Three People in a Botanical Garden. In: *Proc. of the 7th National Conference on Artificial Intelligence*. Morgan Kaufmann, 1988, 846–855.

[Cruz-Neira *et al.* 1992] Cruz-Neira, C. & Sandin, D.J. & DeFanti, T.A. & Kenyon, R.V. & Hart, J.C.: The CAVE: Audio Visual Experience Automatic Virtual Environment. *Communications of the ACM* 35:6, June 1992, 65–72.

[Dankert 1969] Dankert, H.: *Sportsprache und Kommunikation — Untersuchungen zur Struktur der Fußballsprache und zum Stil der Sportberichterstattung*. Tübingen: Tübinger Vereinigung für Volkskunde e.V., 1969.

[Davis 1998] Davis, Erik: *Techgnosis: Myth, Magic, and Mysticism in the Age of Information*. New York: Harmony Books, 1998.

[De Lisi & Cammarano 1996] De Lisi, R., & Cammarano, D.M.: Computer Experience and Gender Differences in Undergratuate Mental Rotation Performance. *Computers in Human Behavior* 12:351–361, 1996.

[Denning 1992] Denning, Peter J.: Educating a New Engineer. *CACM* 5:12, 1992, 83–97.

[Dittmann 2000] Dittmann, Jana: *Digitale Wasserzeichen*. Berlin: Springer, 2000.

[Domik 1994] Domik, Gitta: *Scientific Visualization*. Boulder: University of Colorado,1994.

[Dornes 1993] Dornes, Martin: *Der kompetente Säugling – Die präverbale Entwicklung des Menschen,* Frankfurt/M.: Fischer, 1993.

[Dorsch 1998] Dorsch, Friedrich: *Psychologisches Wörterbuch*, Göttingen: Huber, 1998.

[Dussart 1999] Dussart, Françoise: What an Acrylic Can Mean: On the Meta-Ritualistic Resonances of a Central Desert Painting. In: [Morphy & Smith Boles 1999], 193–218.

[Ehrig & Mahr 1985] Ehrig, Hartmut, & Mahr, B.: *Fundamentals of Algebraic Specification*. Berlin: Springer, 1985.

[Eibl-Eiblesfeld 1984] Eibl-Eiblesfeld, Irenäus: *Die Biologie des menschlichen Verhaltens. Grundriß der Humanethologie.* München: Piper, 1984.

[Fauconnier 1985] Fauconnier, Gilles: *Mental Spaces: Aspects of Meaning Construction in Natural Language*, Cambridge: Cambridge Univ. Press, 1985.

[Fay & Quaiser-Pohl 1999] Fay, E., & Quaiser-Pohl, C.: *Schnitte. Ein Test zur Erfassung des räumlichen Vorstellungsvermögens*. Frankfurt/M: Swets Test Services, 1999.

[Fellmann 1991] Fellmann, Ferdinand: *Symbolischer Pragmatismus. Hermeneutik nach Dilthey*, Reinbek, 1991.

[Fellmann 2000] Fellmann, Ferdinand: Bedeutung als Formproblem – Aspekte einer realistischen Bildsemantik. In: Sachs-Hombach, K. & Rehkämper, K. (eds.): *Vom Realismus der Bilder: Interdisziplinäre Forschungen zur Semantik bildhafter Darstellungsformen*, Magdeburg: Scriptum, 2000, 17–40.

[Files 1996] Files, Craig: Goodman's Rejection of Resemblance. In: *The British Journal of Aesthetics* 10/1996, 398–413.

[Fodor 1976] Fodor, Jerry: The Language of Thought. Hassocks Sussex: Harvester Press, 1976.

[Foley *et al.* 1996] Foley, J.D. & van Dam, A. & Feiner, St.K. & Hughes, J.F.: *Computer Graphics, Principles and Practice*. Addison Wesley, Reading, MA, ^{2}1996.

[Forte & Siliotti 1997] Forte, M. & Siliotti, A. (eds.): *Die neue Archäologie. Virtuelle Reisen in die Vergangenheit*. Bergisch Gladbach: Lübbe, 1997.

[Franke 1987] Franke, Herbert W.: The Expanding Medium: The Future of Computer Art. *Leonardo* 20(4):335–338, 1987.

[Frege 1884] Frege, Gottlob: *Die Grundlagen der Arithmetik. Eine logisch-mathematische Untersuchung über den Begriff der Zahl*. Breslau: Koebner, 1884.

[Frege 1892] Frege, Gottlob: Über Sinn und Bedeutung, in: *Zeitschrift für Philosophie und philosophische Kritik* 100:25–50, 1892.

[Furnas 1986] Furnas, G. W.: Generalized Fisheye Views. In: *Proc. CHI '86 Human Factors in Computing Systems*, ACM SIGGRAPH, 1986, 16–23.

[Gardner & Gardner 1980] Gardner R.A. & Gardner, B.T.: Comparative Psychology and Language Acquisition. In : T.A. Sebeok & Umiker-Sebeok, J. (Eds.): *Speaking of Apes*. New York: Plenum, 1980, 287–330.

[Girmes 1997] Girmes, R. (ed.) : *Studium – Berufsentwicklung – Persönlichkeitsbildung*. Münster: Waxmann, 1997.

[Gombrich 1960] Gombrich, Ernst H.: *Art and Illusion*. Princeton, 1960.

[Gombrich 1984] Gombrich, Ernst: *Bild und Auge. Neue Studien zur Psychologie der bildlichen Darstellung.* Stuttgart: Klett-Cotta, 1984.

[Gonzalez & Woods 2002] Gonzalez, Rafael C. & Woods, R. E.: *Digital Image Processing*. Reading, MA: Addison Wesley, 2002.

[Goodman & Elgin 1988] Goodman, Nelson & Elgin, Catherine Z.: Reconceptions in Philosophy. Indianapolis/Cambridge: Hackett Publishing Co, 1988.

[Goodman 1976] Goodman, Nelson: *Languages of Art. An Approach to a Theory of Symbols*. Indianapolis: Hackett, [1]1968, [2]1976.

[Grau 2003] Grau, Oliver: Charlotte Davies: Osmose. In: *Virtual Art, From Illusion to Immersion*.Cambridge, Ma.: MIT Press, 2003, 193–211.

[Grice 1974] Grice, Herbert Paul: Logic and Conversation. In Cole, V & Morgan, J.L. (eds.): *Syntax and Semantics*, Vol. 3, 41–58. Academic Press, New York, 1974.

[Habermas 1996] Habermas, Jürgen: Technischer Fortschritt und soziale Lebenswelt. In: *Praxis* 1(2): 217–228, 1996.

[Halper *et al.* 2002] Halper, Nick & Schlechtweg, St. & Strothotte, Th.: Creating Non-Photorealistic Images the Designer's Way. In: *Proc. of NPAR 2002* (International Symposium on Non Photorealistic Animation and Rendering), Annecy, 2002, in press.

[Haralick *et al.* 1973] Haralick, R. M. & Shanmugam, K. & Dinstein, I.: Textural Features for Image Classification. *IEEE Transactions on Systems, Man, and Cybernetics* 3(6):610–621, 1973.

[Healey 2000] Healey, C. G.: Building a Perceptual Visualisation Architecture. *Behaviour and Information Technology* 19(5):349–366, 2000.

[Held 1975] Held, J.: Visualisierter Agnostizismus. *Kritische Berichte* 3(5/6):1–7, 1975.

[Hensel 2004FC] Hensel Thomas: Bildwissenschaft als Kunstwissenschaft. Das Paradigma Aby Warburg. To appear in: *Proc. of the International Kolloquium „Bildwissenschaft zwischen Reflexion und Anwendung"* (Magdeburg, September 2003). Köln: Herbert von Harlem, 2004 forthcoming.

[Herskovits 1986] Herskovits, Annette: *Language and Cognition – An Interdisciplinary Study of the Prepositions in English*. Cambridge: Univ. Press, 1986.

[Herzog *et al.* 1990] Herzog, Gerd & Rist, Th. & André, E.: Sprache und Raum: Natürlichsprachlicher Zugang zu visuellen Daten. In Habel, Ch. & Freksa, Ch. (eds.): *Repräsentation und Verarbeitung räumlichen Wissens*. Berlin: Springer, 1990, 207–220.

[Heublein et al. 2002] Heublein Ulrich & Schmelzer R. & Sommer D. & Spangenberg H.: Studienabbruchstudie 2002 – Die Studienabbrecherquote in den Fächergruppen und Studienbereichen der Universitäten und Fachhochschulen. In: *Kurzinformation HIS* A5 2002 (July).

[Hoppe & Lüdicke *1998*] Hoppe Axel & Lüdicke K.: Measuring and Highlighting in Graphics. In. [Strothotte *et al.* 1998], Chapter 7, 121–136.

[HGBRD 2000] Haus der Geschichte der Bundesrepublik Deutschland (ed.): *Bilder, die lügen – X für U*. Bonn: Bouvier-Verlag, ^{2}2000.

[Huber 1997] Huber, Hans-Dieter: Internet. In: Hirner, Rene (ed.): *Vom Holzschnitt zum Internet. Die Kunst und die Geschichte der Bildmedien von 1450 bis heute*. Ostfildern-Ruit: Cantz 1997, 186–189.

[Husserl 1980] Husserl, Edmund: *Logische Untersuchungen*. Critical edition (Husserliana 18, 19/I and II). 1900/01. Reprint: Den Haag: Nijhoff 1975, 1984. Engl. trans. of the 2nd edition by J. N. Findlay, *Logical Investigations*. London: Routledge, 1970.

[Ignatius & Senay 1994] Ignatius, Eve & Senay, H.: A Knowledge-Based System for Visualization Design. *IEEE Computer Graphics and Applications*, 14(6):36–47.

[Institut Mathildenhöhe Darmstadt w/o y.] Institut Mathildenhöhe Darmstadt (ed.): *Museum Künstlerkolonie Darmstadt*. Catalogue, (without year and place).

[Johnson-Laird 1983] Johnson-Laird, P. N.: *Mental Models*. Cambridge: University Press, 1983.

[Kamp & Reyle 1993] Kamp, Hans & Reyle, U.: *From Discourse to Logic: Introduction to Model-theoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory.* Dordrecht: Kluwer, 1993.

[Kant 1781] Kant, Immanuel: *Kritik der reinen Vernunft*. Riga, A: 1781, B: second rev. ed. 1787.

[Kant 1960] Kant, Immanuel: Logik. In : Weischedel, W. (ed.): *Werke in 12 Bänden*. Vol. VI, Wiesbaden, 1960.

[King 2002] King, Mike: *Computers and Modern Art: Digital Art Museum*. Online Article for the Digital Art Museum: http://www.dam.org/essays/king02.htm.

[Kjørup 1978] Kjørup, Soren: Pictorial Speech Acts. *Erkenntnis* 12:55-71, 1978

[Klauck 1994] Klauck, Christoph: *Eine Graphgrammatik zur Repräsentation und Erkennung von Features in CAD/CAM*. DISKI 66. St. Augustin: Infix, 1994, Dissertation.

[Kleene 1967] Kleene, Stephen C.: *Mathematical Logic*. New York: Wiley, 1967.

[Koller 1992] Koller, Dieter: *Detektion, Verfolgung und Klassifikation bewegter Objekte in monokularen Bildfolgen am Beispiel von Verkehrsszenen*. St. Augustin: Infix, 1992.

[Korn 1988] Korn, Axel F.: Toward a symbolic representation of intensity changes in images. *IEEE Transactions on Pattern Analysis and Mashine Intelligence* 10(5):610–625, September 1988.

[Kripke 1972] Kripke, Saul A.: Naming and Necessity. In: Davidson, D. & Harman, G. (eds.): *Semantics of Natural Language*. Dordrecht: Reidel, 1972, 235–355 & 763–769.

[Lakoff & Johnson 1980] Lakoff, George & Johnson, M.: *Metaphors We Live By*. Chicago: University Press, 1980.

[Lau et al. 2001] Lau, L. & Wahyu, D. & Carl, S.: The Power of the Visual Metaphor – or: The Potential Failure of Virtual Reality. In: *CISST'2001*, Las Vegas, 2001.

[Leibniz 1875] Leibniz, Gottfried, W.: Metaphysische Abhandlung. In: Gerhardt C.I. (ed.): *Die philosophischen Schriften*. Vol. VI, Berlin, 1875.

[Lem 1964] Lem, Stanislaw: *Summa technologiae*. Kraków: Wydawnictwo Literackie, 1964.

[Leonard & Goodman 1940] Leonard, H. S. & Goodman, N.: The Calculus of Individuals and its Uses. *Journal of Symbolic Logic* 5(2):45–55 1940.

[Leroi-Gourhan 1984] Leroi-Gourhan, André: *Hand und Wort. Die Evolution von Technik, Sprache und Kunst*, Frankfurt/M.: Suhrkamp, 1984.

[Lettvin *et al.* 1959] Lettvin, Jerome Y. & Maturana, H. & McCulloch, W. & Pitts, W.: What the Frog's Eye Tells the Frog's Brain. In: *Proc. of the Institute of Radio Engineers* 47/11, 1959.

[Levin *et al.* 1987] Levin, J.R. & Anglin, G.J. & Carney, R.N.: On Empirically Validating Functions of Pictures in Prose. In: Willows, D. A. & Houghton, H. A. (Eds): *The Psychology of Illustration*, Vol. 1, Berlin: Springer, 51–85, 1987

[Levinson 1983] Levinson, Stephen C.: Pragmatics. Cambridge Univ. Press, 1983.

[Lévi-Strauss 1992] Lévi-Strauss, Claude: *Tristes Tropiques.* Penguin Books, 1992.

[Lewis 1986] Lewis, David: *On the Plurality of Worlds*. Oxford, 1986.

[Lindsay & Norman 1977] Lindsay, Peter H. & Norman, D. A.: *Human Information Processing*. New York: Academic Press, 1977.

[Lock 1993] Lock, Andrew: Human Language Development and Objecvt Manipulation: Their Relation in Ontogeny and Its Possible Relevance for Phylogenetic Questions. In: Gibson, K. & Ingold, T. (Eds.): *Tools, Language and Cognition in Human Evolution*. Cambridge: Cambridge Univ. Press, 1993, 279–299.

[Long et al. 2000] Long, Hui & Leow, K. W. & Chua, F.K. & Kee, F. : Perceptual Texture Space for Content-Based Image Retrieval. In: *Proc. MMM*, 167–180, 2000.

[Lopes 1996] Lopes, Dominic: *Understanding Pictures.* Oxford: Clarendon Press, 1996.

[Lubich 1995] Lubich, H.P.: *Towards a CSCW Framework for Scientific Cooperation in Europe*. Lecture Notes in Computer Science 889. Berlin, Heidelberg, New York: Springer-Verlag, 1995.

[Maier 1996] Maier, P. H. : Geschlechtsspezifische Differenzen im räumlichen Vorstellungsvermögen. *Psychologie in Erziehung und Unterricht* 43:245–265, 1996.

[Mann & Thompson 1987] Thompson, Sandra A. & Mann, W. C.: Rhetorical Structure Theory: A Framework for the Analysis of Texts. In: Duranti, A. & Schieffelin, B. (eds.), *International Pragmaties Association Papers in Pragmatics* 1, 79–105.

[Marr 1982] Marr, David: *Vision: A Computational Investigation into Human Representation and Processing of Visual Information*. New York: Freeman & Co., 1982.

[Masuch 2001] Masuch, Maic: *Nicht-photorealistische Visualisierungen: Von Bildern zu Animationen*. Aachen: Shaker, 2002.

[M^{C}Cloud 1993] M^{C}Cloud, Scott: *Understanding Comics – The Invisible Art.* Northampton (Ma): Kitchen Sink Press, 1993.

[Mead 1932] Mead, George Herbert: The Physical Thing. In: Mead, G.H.: *The Philosophy of the Present*. (ed.: A.E. Murphy), Chicago: Open Court, 1932, 119–139.

[Mead 1968] Mead, George Herbert: *Geist, Identität und Gesellschaft aus der Sicht des Sozialbehaviorismus*, Frankfurt a. M.: Suhrkamp, 1968. (^{1}1934: *Mind, Self, and Society*, Chicago: Univ. Press)

[Meier-Graefe 1901] Meier-Graefe, Julius: Darmstadt. *Die Zukunft* 35:478–486, 1901.

[Metzger 1966] Metzger, Wolfgang: Figurale Wahrnehmung. In: Metzger W. (Ed.): *Handbuch der Psycholohgie*, Vol. 1, Göttingen, Hogrefe, 1966.

[Mitchell 1992] Mitchell, W. J. T. : The Pictorial Turn. *Art Forum* 1992(March):89–95.

[Mitchell 1994] Mitchell, W. J. T.: *Picture Theory*. Chicago: University Press, 1994.

[Mohr 1976] Mohr, Manfred: Statement in Leavitt, Ruth (Ed.): *Artist and Computer*. San Jose: Harmony Press, 1976, p. 96.

[Morphy & Smith Boles 1999] Morphy, Howard & Smith B. M. (eds.): *Art from the Land: Dialogues with the Kluge-Ruhe Collection of Australian Aboriginal Art.* Charlottesville: University of Virginia, 1999.

[Nake 1996] Nake, Frieder: Einmaliges und Beliebiges. Künstliche – Kunst im Strom der Zeit. In: Erdmann, J.W. & Rückriem, G. & Wolf, E. (eds.): *Kunst, Kultur und Bildung im Computerzeitalter*. Berlin: Hochschule der Künste, 1996. 41–58.

[Noble & Davidson, 1996] Noble, William & Davidson, I.: *Human Evolution, Language, and Mind – A Psychological and Archeological Inquiry*. Cambridge: Univ. Press, 1996.

[Newcombe & Dubas 1992] Newcombe, N., & Dubas, J. S.: A longitudinal study of predictors of spatial ability in adolescent females. *Child Development* 63: 37–46, 1992.

[Parunak 1996] Parunak, H. V. D.: Visualizing Agent Conversations: Using Enhanced Dooley Graphs for Agent Design and Analysis. In: *Proceedings of the 2nd International Conference on Multi-Agent Systems (ICMAS '96)*, Kyoto, Japan, 275–282, 1996.

[Peirce 1931ff] Peirce, Charles Sanders: *Collected Papers* (Vol 1 – 6) (Eds. : Hartshorne Ch. & Weis P.), Cambridge (Mass.): Harvard Univ. Press. 1931–1935.

[Peirce 1903] Peirce, Charles Sanders: A Syllabus of Certain Topics of Logic. 1903. Reprinted in: *The Essential Peirce. Selected Philosophical Writings*. Vol. 2 (1893-1913). Bloomington and Indianapolis: Indiana University Press, 1998.

[Pineda 2003FC] Pineda, Luis A.: Multimodal generation, spatial language and illustration. To appear in: *Proc. of the ECAI 2003*. http://leibniz.iimas.unam.mx/~luis/papers/ilustra-paper1.pdf

[Plessner 1928] Plessner, Helmuth: *Die Stufen des Organischen und der Mensch*. Berlin: W. de Gruyter, ^{3}1975, (^{1}1928).

[Plinius 1977] Plinius Secundus, Gaius: *Naturalis Historiae Libri*, Liber XXXV, Düsseldorf/Zürich: Artemis & Winkler, 1997.

[Plümacher 1999] Plümacher Martina: Wohlgeformtheitsbedingungen für Bilder?. In: Sachs-Hombach, K. & Rehkämper, K.: *Bildgrammatik: Interdisziplinäre Forschungen zur Syntax bildlicher Darstellungsformen.* Magdeburg: Scriptum, 1999, 47–56.

[Postman 1985] Postman, Neil: *Wir amüsieren uns zu Tode*. Frankfurt am Main: Fischer 1985.

[Preim 1998] Preim Bernhard: *Entwicklung interaktiver Systeme*. Berlin: Springer, 1998.

[Preim & Hoppe 1998] Preim, Bernhard & Hoppe, A.: Enrichment and Reuse of Geometric Models. In: [Strothotte et.al. 1998], Chapter 3, 45–62.

[Pross 1972] Pross, Harry: *Medienforschung.* Darmstadt: Habel, 1972.

[Prusinkiewicz & Lindenmayer 1990] Prusinkiewicz, Przemlyslaw & Lindenmayer, Aristid: *The Algorithmic Beauty Of Plants*, Springer Verlag, 1990.

[Quaiser-Pohl 2001] Quaiser-Pohl, C. : Mentales Rotieren und kognitive Landkarten – Über geschlechtsspezifische Unterschiede im räumlichen Denken. In Helfrich, H. (ed.): *Frauen denken anders*. Opladen: Leske+Budrich, 2001.

[Quaiser-Pohl *et al.* 2001] Quaiser-Pohl, Claudia & Lehmann, W. & Schirra, J.R.J. : Sind Studentinnen der Computervisualistik besonders gut in der Raumvorstellung? Psychologische Aspekte bei der Wahl eines Studienfachs. *Fiff-Kommunikation* 18(3): 42–46 (3/2001).

[Rackham 2000] Rackham, Melinda: Empyrean – Soft Skinned Space. In: Hollands, Robin (ed.): *Proceedings of the 7th UK VR-SIG Conference*, 135–140.

[Randow 1995] Randow, Gero v.: Die neue Macht des Auges. *Die Zeit,* July 7, 1995, 49f.

[Rehkämper 2002] Rehkämper, Klaus: *Bilder, Ähnlichkeit und Perspektive. Auf dem Weg zu einer neuen Theorie der bildhaften Repräsentation*. Wiesbaden: DUV, 2002.

[Reichle 2002] Reichle, Ingeborg: TECHNOSPHERE: Körper und Kommunikation im Cyberspace. In. Sachs-Hombach, K. (ed.): *Bildhandeln*. Magdeburg: Scriptum, 2002, 193–204.

[Rist 1996] Rist, Thomas: *Wissensbasierte Verfahren für den automatischen Entwurf von Gebrauchsgraphik in der technischen Dokumentation*. DISKI 139. St. Augustin: Infix, 1996.

[Ros 1989 & 1990] Ros, Arno: *Begründung und Begriff*. Hamburg: Meiner, 3 Vol., 1989/1990.

[Ros 1999] Ros, Arno: Was ist Philosophie? In: Richard Raatzsch (ed.): *Philosophieren über Philosophie*. Leipzig: Leipziger Universitätsverlag, 1999, 36–58.

[Ros 2005] Ros, Arno: *Materie und Geist. Eine philosophische Untersuchung*. Paderborn: Mentis, 2005.

[Rowling 1998] Rowling, Joanne K.: *Harry Potter and the Chamber of Secrets*. New York: Arthur A. Levine/Scholastic Press.

[Rumsey 2001] Rumsey, Alan: Tracks, Traces, and Links to Land in Aboriginal Australia, New Guinea, and Beyond. In: Rumsey, A. & Weiner, J.F. (eds.): *Emplaced Myth – Space, Narrative, and Knowledge in Aboriginal Australia and Papua New Guinea*. Honolulu: Univ. Hawaii Press, 19–42.

[Sachs-Hombach & Rehkämper 1999] Sachs-Hombach, K. & Rehkämper, K.: Aspekte und Probleme der bildwissenschaftlichen Forschung – Eine Standortbestimmung. In: Sachs-Hombach, K. & Rehkämper, K.: *Bildgrammatik: Interdisziplinäre Forschungen zur Syntax bildlicher Darstellungsformen.* Magdeburg: Scriptum, 1999, 9–20.

[Sachs-Hombach & Schirra 1999] Sachs-Hombach, Klaus & Schirra, J.R.J.: Zur politischen Instrumentalisierbarkeit bildhafter Repräsentationen. Philosophische und psychologische Aspekte

der Bildkommunikation. In: Hofmann, W. (ed.): *Die Sichtbarkeit der Macht – Theoretische und empirische Untersuchungen zur visuellen Politik*. Baden-Baden: Nomos, 1999, 28–39.

[Sachs-Hombach & Schirra 2002] Von der interdisziplinären Grundlagenforschung zur computervisualistischen Anwendung. Die Magdeburger Bemühungen um eine allgemeine Wissenschaft vom Bild. *Magdeburger Wissenschaftsjournal* 1/2002, 27–38.

[Sachs-Hombach 2001] Sachs-Hombach, Klaus: Bildbegriff und Bildwissenschaft. In: Gerhardus, D. & Rompza, S. (eds.): *Kunst – Gestaltung – Design*, Vol. 8. Saarbrücken: St. Johann, 2001.

[Sachs-Hombach 2002] Sachs-Hombach, Klaus: *Elemente einer allgemeinen Bildwissenschaft*. Habilitationsschrift, FGSE, Univ. Magdeburg 2002. (revised version: Köln: Herbert von Harlem, 2003)

[Sachs-Hombach 2003] Sachs-Hombach, Klaus: Resemblance Reconceived. In: Hecht, H. & Schwartz, B. & Atherton, M. (eds.): *Looking into Pictures. An Interdisciplinary Approach to Pictorial Space*. Cambridge: MIT Press, 2003, 167-177.

[Sachs-Hombach 2005] Sachs-Hombach, Klaus: *Bildwissenschaft. Disziplinen, Themen, Methoden*. Frankfurt/M.: Suhrkamp, 2005.

[Sacks 1986] Sacks, Oliver: *The Man Who Mistook His Wife for a Hat*. New York: Picador, 1986.

[Salvator et al. 2001] Salvador, Elena & Cavallaro, A. & Ebrahimi, T.: Shadow Identification and Classification Using Invariant Color Models. In: *Proc. ICASSP2001, 2001,* Salt Lake City, UT, Vol 3: 1545–1548, 2001.

[Sarkar & Brown 1992] Sarkar, Manojit & Brown, M. H.: Graphical Fisheye Views of Graphs. In: *Human CHI'92 Conference Proceedings*, 83–91, 1992.

[Schäfer 1901] Schäfer, Wilhelm: 'Ein Dokument deutsche Kunst'. Die Ausstellung der Künstler-Kolonie Darmstadt 1901. *Die Rheinlande* 2:38–39, 1901.

[Scheffler 1902] Scheffler, Karl: Das Haus Behrens. In: *Dekorative Kunst* 9:1–48, 1902.

[Schinzel *et al.* 1999] Schinzel, Britta & Kleinn, K. & Wegerle, A. & Zimmer, Chr.: Das Studium der Informatik: Studiensituation von Studentinnen und Studenten. *Informatik-Spektrum* 22(1):13–23, 1999.

[Schirra 1999] Schirra, J.R.J.: Computational Visualistics: Bridging the Two Cultures in a Multimedia Degree Programme. In: Pudlowski, Z. J. (ed.): *Proc. of the 2nd Asia-Pacific Forum on Engineering & Technology Education, Sydney*. Melbourne: UICEE, 47–51, 1999.

[Schirra 1997] Schirra, J.R.J.: Referenzsemantik in der Hörermodellierung. *Kognitionswissenschaft* 6(4):177–196, 1997.

[Schirra 1995] Schirra, J.R.J.: Understanding Radio Broadcasts On Soccer: The Concept »Mental Image« and Its Use in Spatial Reasoning. In: K. Sachs-Hombach, (ed.): *Bilder im Geiste: Zur kognitiven und erkenntnistheoretischen Funktion piktorialer Repräsentationen*. Rodopi, Amsterdam, 1995, 107–136.

[Schirra 1994] Schirra, J.R.J.: *Bildbeschreibung als Verbindung von visuellem und sprachlichem Raum — Eine interdisziplinäre Untersuchung von Bildvorstellungen in einem Hörermodell*. St. Augustin: infix, 1994.

[Schlechtweg & Raab 1998] Schlechtweg, Stefan & Raab, A.: Rendering Line Drawings for Illustrative Purposes. In. [Strothotte *et al.* 1998], Chapter 4, 63–89.

[Schlechtweg & Wagner *1998*] Schlechtweg, Stefan & Wagner, Hubert: Interactive Medical Illustrations. In. [Strothotte *et al.* 1998], Chapter 17, 295–311.

[Schlechtweg & Strothotte 1996] Schlechtweg, Stefan & Strothotte, Th.: Rendering Line-Drawings with Limited Resources. In: *Proc. GRAPHICON '96*, *6th International Conference and Exhibition on Computer Graphics and Visualization in Russia*, vol. 2, 131–137, 1996.

[Scholz 1991] Scholz, Oliver: *Bild, Darstellung, Zeichen – Philosophische Theorien bildhafter Darstellung*, Freiburg/München: Alber, 1991.

[Schwan 2001] Schwan, Stephan: *Filmverstehen und Alltagserfahrung. Grundzüge einer kognitiven Psychologie des Mediums Film*. Wiesbaden: DUV, 2001.

[Searle 1969] Searle, John: *Speech Acts*. Cambridge: Univ. Press, 1969.

[Smith 2002] Smith, J. H.: *The Architectures of Trust – Supporting Cooperation in the Computer-Supported Community*. Master's Thesis, Department of Film and Media Studies, University of Copenhagen, 2002:
http://www.gamasutra.com/education/theses/20020410/smith.pdf.

[Smith 1996] Smith, Barry: Mereotopology: A Theory of Parts and Boundaries. *Data and Knowledge Engineering*, 20:287–303, 1996.

[Snow 1959] Snow, C.P.: *The Two Cultures and the Scientific Revolution*. Rede Lecture, Cambridge: Cambridge Univ. Press, 1959.

[Steinbrenner 1999] Steinbrenner Jakob: Zitatzeit – oder Füßchen der Gänse überall – oder worauf Zitate Bezug nehmen. In: Sachs-Hombach, K. & Rehkämper, K. (eds.): *Bildgrammatik: Interdisziplinäre Forschungen zur Syntax bildlicher Darstellungsformen*. Magdeburg: Scriptum, 1999, 135–144.

[Steller 1992] Steller, Erwin: *Computer und Kunst. Programmierte Gestaltung:. Wurzeln und Tendenzen neuer Ästhetiken*. Mannheim: Bibliographisches Institut Verlag, 1992.

[Strassmann 1986] Strassmann, S.: Hairy Brushes. In: *Proceedings of SIGGRAPH'86*, 225–232, 1986

[Strothotte 1989] Strothotte, Thomas: *Interaktive und wissensbasierte Methoden zur integrierten Bild- und Sprachkommunikation*. Habilitationsschrift, Stuttgart, 1989.

[Strothotte 1998] Strothotte, Thomas: Integrating Spatial and Nonspatial Data. In: [Strothotte *et al.* 1998], Chapter 23, 403–418.

[Strothotte & Strothotte 1997] Strothotte, Christine & Strothotte, Th.: *Seeing Between the Pixel – Pictures in Interactive Systems*. Berlin: Springer, 1997.

[Strothotte *et al.* 1998] Strothotte, Th., *et 18 al.*: *Computer Visualization – Graphics, Abstraction, and Interactivity*. Springer, Heidelberg, 1998

[Strothotte & Schlechtweg 2002] Strothotte, Thomas & Schlechtweg, St.: *Non-Photorealistic Computer Graphics. Modeling, Rendering, and Animation*. Morgan Kaufmann, San Francisco, 2002.

[Sung 1988] Sung C.-K.: Extraktion von typischen und komplexen Vorgängen aus einer langen Bildfolge einer Verkehrsszene. In: Bunke H. & Kübler, O. & Stucki, P.: (eds.): *Musterer-kennung 88*, Berlin: Springer, 1988, 90–96.

[Sylvester 1997] Sylvester, David: *About Modern Art: Critical Essays, 1948–1997*, New York: Henry Holt, 1997.

[Tinbergen 1951]Tinbergen, Niklaas: *The Study of Instinct*. London: Oxford University Press, 1951.

[Tugendhat 1982] Tugendhat, E.: *Traditional and Analytic Philosophy*. Cambridge: Cambridge University Press, 1982. (citation from [1]1976: *Vorlesungen zur Einführung in die sprachanalytische Philosophie*, Frankfurt/M.: Suhrkamp).

[Turoff et al. 1999] Turoff, M. & Hiltz, S.R. & Bieber, S.R. & Fjermestad, J. & Rana, A.: Collaborative Discourse Structures in Computer Mediated Group Communications. *JCMC* 4(4).

[Uexküll 1909] Uexküll, Jacob v.: *Umwelt und Innenwelt der Tiere*, Berlin: Springer, 1909.

[Ulmer 2001] Ulmer, R.: Biomorphismus. In: Buchholz K., Latocha, R., Peckmann, H. & Wolbert K. (eds.): *Die Lebensreform – Entwürfe und Neugestaltung von Leben und Kunst um 1900*. 2 Vol.s, Darmstadt: Haeuser & Institut Mathildenhöhe, Vol. 2, 455–470.

[van Deemter 1998] Van Deemter, K.: Retrieving Pictures for Document Generation. In *Proc. of 14th Workshop on Language Technology*, University of Twente, The Netherlands, 1998, 117–128.

[Vandenberg & Kuse 1978] Vandenberg, S. G. & Kuse, A.R.: Mental Rotations. A Group Test of Three-Dimensional Spatial Visualization. *Perceptual and Motor Skills* 47:599–604, 1978.

[Vieu 1991] Vieu, V.: *Sémantique des relations spatiales et inférences spatio-temporelles: Une contribution à l'étude des structures formelles de l'espace en Langage Naturel*. Dissertation, Institut de Recherches en Informatique de Toulouse, Univ. Paul Sabatier, Toulouse, 1991.

[Wahlster 1991] Wahlster, Wolfgang: User and Discourse Models for Multimodal Communication. In: Sullivan, J. W. & Tyler, Sh. W. (eds.): *Intelligent User Interfaces*. ACM Press, Frontier Series. 45–67, 1991.

[Watson 1999] Watson, Christine: Touching the Land: Towards an Aesthetic of Balgo Contemporary Painting. In: [Morphy & Smith Boles 1999], 163–192.

[Watt & Policarpo 2001] Watt, Alan & Policarpo, F.: *3D Games: Real-Time Rendering and Software Technology*. New York: Addison Wesley, 2001.

[Whitehead 1929] Whitehead, Alfred North: *Process and Reality*. New York: The MacMillan Company, 1929.

[Winograd & Flores 1986] Winograd, T. & Flores, F.: *Understanding Computers and Cognition.* Addison-Wesley, Boston, 1986.

[Wittgenstein 1984] Wittgenstein, Ludwig: *Bemerkungen über die Farben*. In: Vol. 8 of *Werkausgabe: Über Gewißheit.* Frankfurt: Suhrkamp, 1984

[Wittgenstein 1953] Wittgenstein,Ludwig: *Philosophical Investigations*. New York: MacMillan, 1953.

[Wong & Bergeron 1995] Wong, Pak Chung & Bergeron, R. D. : 30 Years of Multidimensional Multivariate Scientific Visualization. In: Nielson, G. M. & Hagan, H. & Muller, H. (eds.): *Visualization – Overviews, Methodologies and Techniques*. IEEE Computer Society Press, 3–33, 1997.

[Wu & Chen 1992] Wu, C.-M. & Chen, Y.-C.: Statistical Feature Matrix for Texture Analysis. *CVGIP: Graphical Models and Image Processing*, 54(5): 407–419.

All webpages referred to have been verified in October 2003.

B List of Figures

C List of Tables

D Overview in German

1 Eine Lageanalyse: Das Kapitel legt Motivation und gesetzte Ziele der Arbeit dar.

1.1 Bilder nehmen in der Gegenwart eine zunehmend wichtige Rolle ein, was allerdings gemeinhin ambivalent aufgenommen wird. Der technische Fortschritt hat einen zentralen Anteil an dieser Entwicklung. Eine allgemeine Bildwissenschaft, die der gesteigerten Wichtigkeit Rechnung trägt und nicht nur Einzelaspekte betrachtet, bildet sich allerdings gerade erst heraus.

1.2 Die zunehmende Wichtigkeit der Bilder folgt einem allgemeineren Trend der westlichen Gesellschaft hin zur Informationsgesellschaft: wichtiger Teil davon ist die Reflexion der Medien.

1.3 Computer und Computernetze übernehmen in der Informationsgesellschaft eine tragende Funktion: dies zeigt sich insbesondere auch darin, daß sie den Bildbegriff entscheidend erweitert haben um das „interaktive Bild". „Computervisualistik" wird als neue Bezeichnung für eine Unterdisziplin der Informatik eingeführt. In ihr wird versucht, systematisch alle computerrelevanten Aspekte einer Wissenschaft vom Bild anzugehen.

1.4 Dabei sollte sich die Computervisualistin/der Computervisualist als Vertreter einer modernen, nicht-technokratischen Auffassung des Ingenieurberufs sehen, wie sie in den letzten Jahren im Nachklang der von SNOW ausgelösten Zwei-Kulturen-Debatte vielfach beschworen wurde. Eine Grundlegung der Computervisualistik hat dem Rechnung zu tragen.

1.5 Damit ist die Ausgangssituation umschrieben: Vor diesem Hintergrund ist das Ziel der Arbeit, eine fundierte Übersicht über die Computervisualistik und ihre Grundlagen zu geben, wie sie bislang noch nicht vorlag. Der Abschnitt schließt mit einer Übersicht über die folgenden Kapitel.

2 Computervisualistik als Fach: Das Kapitel thematisiert die methodologischen Grundlagen und die interdisziplinäre Einordnung der Computervisualistik.

2.1 Als Kernthemen der Informatik werden »(abstrakte) Datenstruktur« und »Implementierung« herausgearbeitet: dabei steht die Informatik selbst im Einflußbereich zweier Methodologien, insofern sie sich sowohl als Strukturwissenschaft wie als Ingenieurwissenschaft versteht. Beide Herangehensweisen profitieren von einer argumentationstheoretischen Analogie, bei der Datenstrukturen als formalisierte Begriffsfelder verstanden werden und Implementierung als eine sehr komplexe Form des Argumentierens.

2.2 Eine allgemeine Bildwissenschaft als die zweite Wurzeldisziplin der Computervisualistik bildet sich gerade erst aus; d.h.: bislang gab es eher die vielen Bildwissenschaften als die eine Bildwissenschaft. Dieser Abschnitt umreißt daher nur kurz, welche Disziplinen vor allem dazu beitragen und welche Schritte zu einer allgemeinen Visualistik unternommen wurden (eine genauere Auseinandersetzung mit dem zugrundeliegenden Kernthema „Bild" erfolgt in Kapitel 3).

2.3 Damit ergibt sich für die Computervisualistik, daß sie sich mit den Mitteln der Informatik dem Thema widmet, das für die Bildwissenschaft konstituierend ist. In der informatischen Form wird daraus ein (abstrakter) Datentyp »Bild« und die dazugehörigen Operationen. Zwei mögliche Untergliederungen des Aufgabenfeldes werden dargelegt; eine entspricht den bislang entwickelten Teilbereichen der Informatik, die sich mit Aspekten von Bildern beschäftigen und daher als Teil der umfassenderen Computervisualistik aufzufassen sind. Die zweite weist unterschiedliche Methoden auf, die in allen Teilbereichen gleichermaßen verwendet werden.

3 Was genau das Thema der Computervisualistik ist, was nämlich ein Bild ist, das ist etwas, was die allgemeine Bildwissenschaft klärt: das Kapitel liefert eine dreifache Übersicht: anhand von Beispielen (3.1), anhand einer Definition (3.2), und anhand einer logisch-genetischen Skizze zur Motivation der Definition (3.3 – 3.5)

3.1 Um den Umfang des Bildbegriffs abzustecken, wird eine Reihe von Beispielen mit eher randständigen Bildformen gegeben.

3.2 Bei den gegenwärtigen Bemühungen um eine allgemeine Bildwissenschaft steht ein bestimmter Vorschlag zum Verständnis dessen, was den Begriff »Bild« bestimmt, im Mittelpunkt: Dieser Vorschlag, Bilder als „wahrnehmungsnahe Zeichen" zu charakterisieren, führt zwei konkurrierende Traditionslinien zusammen, nämlich die semiotischen und die phänomenologischen Bildtheorien.

3.2.1 Eine genauere Erläuterung dieser Begriffsbestimmung setzt die Klärung des Begriffs eines Zeichens im allgemeinen voraus: damit beschäftigt sich die Semiotik. Zentraler Punkt ist, daß Zeichen nur als Teile von Zeichenhandlungen auftreten. Zeichenhandlungen sind recht komplexe Formen interaktiven Verhaltens. Daher können Zeichen unter sehr vielen verschiedenen Aspekten betrachtet werden. Hervorzuheben sind hierbei insbesondere Pragmatik, Semantik und Syntax, sowie Repräsentation, Ausdruck und Appell.

3.2.2 Das Spezifikum bildhafter Zeichen, wahrnehmungsnah zu sein, hängt mit der umgangssprachlichen Auffassung zusammen, daß Bilder dem Abgebildeten ähnlich seien. Wahrnehmungspsychologische Aspekte müssen also eine gegenüber anderen Zeichenarten herausgehobene Rolle spielen.

3.2.3 An dieser Stelle wird es wichtig, die Gemeinsamkeiten (und Unterschiede) zu einer im Zusammenhang mit Bildern häufig verwendeten semiotischen Einteilung von Zeichen zu betrachten: Anzeichen, Ikon und Symbol (nach PEIRCE). Eine eindeutige Zuordnung von Bild zu Ikon ist nicht sinnvoll.

Der Bildbegriff wurde also, verkürzt gesagt, in einer Zusammenführung der Begriffe „Ähnlichkeit" und „Zeichenverwendung" bestimmt. Die folgenden Unterkapitel skizzieren, wie zwei zunächst unabhängige logisch-genetische Herleitungen der beiden Begriffe ineinander greifen (können), um so den vorgestellten Bildbegriff zu ermöglichen. Dieser etwas ausführliche Exkurs, der hier erstmals in vollem Umfang veröffentlicht ist, ist motiviert darin, daß eine solche logisch-genetische Untersuchung weitere Randbedingungen des Bildbegriffs motiviert.

3.3 Der Zusammenhang zwischen Bild und Abgebildetem, also die bildhafte Repräsentationsbeziehung, wird als eines der zentralen Elemente jeder Bildtheorie untersucht. Das führt zu der Frage, was eigentlich Ähnlichkeit heißt, bzw. unter welchen Voraussetzungen wir jemandem zuschreiben, daß er einen Fall von Ähnlichkeit erkannt habe.

3.3.1 Der naive Versuch einer Ähnlichkeitstheorie wird umrissen und zugunsten eines handlungstheoretischen Konzepts abgewiesen. Diese handlungstheoretische Alternative wird in den Abschnitten 3.3 und 3.4 skizziert.

3.3.2 Ähnlichkeit zu erkennen ist ein Begriff, der auf bereits recht komplexen Verhaltensfähigkeiten aufsetzt. Sind die Fähigkeiten eines Wesens, sich zu verhalten, nicht komplex genug (etwa nur Reflexe), dann ist es nicht einmal sinnvoll, von „wahrnehmen" oder „sich täuschen" zu reden.

3.3.3 Ein sinnvoller Wahrnehmungsbegriff setzt eine erste Form der Objektkonstitution voraus. Der Begriff der von solchen Wesen wahrgenommenen (Prä-) Objekte ist allerdings nicht mit dem zu verwechseln, was wir Menschen normalerweise einen (materiellen) Gegenstand nennen würden. Ein solches Wesen kann sich täuschen, z.B. weil etwas et-

was anderem sehr ähnlich sieht; aber es kann diese Ähnlichkeit selbst nicht wieder erkennen. Es werden daher zwei Typen von Ähnlichkeit als handlungstheoretisch verschieden eingeführt (mit Index *alpha* und *beta*). Der komplexere und für den Bildbegriff relevante Begriff der Ähnlichkeit hängt damit zusammen, sich zugleich des anwesenden Täuschenden *und* des abwesenden Verwechselten bewußt zu sein.

3.4 Der Übergang zu dem komplexeren Begriff von Ähnlichkeit (*beta*), den wir normalerweise im Sinn haben, hängt eng mit dem Verhältnis zwischen Bild und Sprache zusammen, das in diesem Unterkapitel untersucht wird.

3.4.1 Es wird zunächst kurz zusammengefaßt, was als das Charakteristische an Sprache gilt. Die aus der logisch-semantischen Propädeutik bekannte Aufteilung der kommunikativen Funktionen der Aussage in die ungesättigten Teilfunktionen Nomination und Prädikation bildet den Ausgangspunkt und führt zur Feststellung, daß Aussagen stets kontext-relativ aber nicht kontext-abhängig verwendet werden. Der aktuelle Verhaltenskontext ist nur eine Möglichkeit.

3.4.2 Dieser Befund wird anhand einer Gegenkonzeption verdeutlicht: Man kann sich leicht eine kontext-abhängige Form der Kommunikation vorstellen. Sie entspricht der in 3.3.3 beschriebenen Verhaltenskomplexität. Wesen, die nur über solche Kommunikationsarten verfügen, kommunizieren lediglich über das, was für sie aktuell da ist. Ähnlichkeit im anspruchsvollen Sinn zu erkennen beruht aber vor allem darauf, zu erkennen, was nicht aktuell da ist.

3.4.3 Die Kontextrelativität aussageartiger Zeichenhandlungen setzt voraus, daß ein Kontext mehr oder weniger explizit angegeben wird, in dem die referentielle Verankerung der anderen Teilhandlungen (insbesondere die empirische Verifikation der Prädikation) stattfinden kann: eine entsprechende Teilhandlung wird unter dem Namen „Kontextbilder“ eingeführt. Damit verbunden ist die Fähigkeit, nicht-anwesende Verhaltenssituationen zu evozieren.

3.4.4 Der Begriff der Gegenstände, wie wir ihn normalerweise verwenden, hängt entscheidend daran, daß wir solche Gegenstände als Elemente vieler Kontexte auffassen: es sind Sortale, die trotz verschiedener Gegebenheitsweisen als identisch verstanden werden. Eine konstituierende Rolle für Sortale spielen einerseits Gestalt-Aspekte, die vor allem instantan und geometrisch-visueller Natur sind, und abstrakte Teil-Ganzes-Beziehungen, die den Zusammenhang zwischen den verschiedenen Gegebenheitsweisen vermitteln: das Verhältnis bestimmt die zweite Objektkonstitution. Erst im Zusammenspiel zwischen geometrischen Begriffen und den Sortalen kann sich zudem die Vorstellung vom reinen (leeren) Raum ausbilden, in dem sich Objekte befinden und bewegen, und der durch Koordinaten bestimmt ist.

3.4.5 Insgesamt folgt, daß Bilder die Fähigkeit zu aussage-artiger Sprache voraussetzen. Ihre Grundfunktion wird als die eines Kontextbilders begriffen. Eine direkte Zuordnung der Bildfunktion zu der der Aussage oder einer ihrer anderen Teilfunktionen ist als davon abgeleitet zu verstehen. Bilder sind besondere Kontextbilder, da sie, im Gegensatz zu den anderen meist sprachlichen Formen der Kontextbildung, einen nicht-anwesenden Kontext derart mit dem aktuellen Verhaltenskontext verschmelzen, daß zumindest teilweise die zugehörigen senso-motorischen Verifikationsroutinen zum Einsatz kommen können (im Sinne einer kontrollierten Verwechslung).

Damit wurde die handlungstheoretische Konzeption von Ähnlichkeit mit der Zeichenverwendung zusammengeführt.

3.5 Was genau es heißt, daß eine kontrollierte Verwechslung der Bildverwendung zugrunde liegt, hängt vom Verhältnis zwischen Bild und Bildbenutzer ab. Zudem ist zu klären, wie nicht-repräsentationale Arten von Bildern in die Theorie passen.

3.5.1 Verschiedene Reflexionsmodi beim Umgang mit Bildern sind zu unterscheiden. Der dezeptive Modus im Verhalten einem Bild(träger) gegenüber entspricht der unreflektierten Form von Ähnlichkeit (alpha), also einer unmittelbaren Verwechslung ohne Zeichengebrauch. Der symbolische Modus faßt die Verwendung als Zeichen ganz allgemein zusammen: der Zeichenträger dient zur Evokation eines Abwesenden. Das spezifisch Bildhafte ist dabei noch nicht berührt. Der immersive Modus besteht aus einer Verkoppelung der beiden anderen Modi: das Bild löst unmittelbar dezeptive Reaktionen beim Bildnutzer aus, die dieser aber erkennt, da er das Bild als Zeichen verwendet. Die spontanen Reaktionen werden dadurch zum Teil unterdrückt, zum Teil (im Rahmen der Zeichenhandlung) kontrolliert aufrechterhalten. Als Verwendung höherer Ordnung läßt sich schließlich der reflexive Modus unterscheiden, bei dem ein Bild Aspekte der Bildkommunikation exemplifiziert.

3.5.2 Das Wechselspiel der Reflexionsmodi bei der Bildverwendung muß insbesondere dem Produzenten in vollem Maße klar sein. Auch wenn das Produkt mit der Intention angefertigt wurde, nur den dezeptiven Modus auszulösen (etwa bei *trompe l'œil*-Bildern, sofern sie nicht reflexiv gemeint sind). Eine Bestimmung von Realismus und Naturalismus (wie er im weiteren verwendet wird) geht der Frage nach der Wahrheit von Bildern voraus: statt Wahrheit sollte allerdings besser über Authentizität von Bildhandlungen gesprochen werden, da Bilder nicht analog zu Aussagen verstanden werden.

3.5.3 Zugleich stellt sich die Frage, wie diese semiotische Auffassung bei den sogenannten „natürlichen Bildern", etwa einem Spiegelbild, ausfällt. Es wird vorgeschlagen, sie als bildliche Selbstgespräche zu verstehen, wobei die Frage nach der Authentizität auch hier eine befriedigende Antwort erlaubt: jemand, der etwas als „natürliches Bild" in einem solchen Selbstgespräch verwendet, obwohl er weiß, daß die Entstehungsbedingungen fehlerhaft waren, handelt nicht authentisch. Um Authentizität erzeugen oder einschätzen zu können, müssen sich die Partner der Zeichenhandlung ihrer Rollen wechselseitig bewußt sein.

3.5.4 Mithilfe verschiedener rhetorischer Ableitungen können auch Abstraktionen und Strukturbilder durch die in den vorangegangenen Abschnitten skizzierte Bildtheorie behandelt werden. Abstraktionen in repräsentationalen Bildern können als graphisches Äquivalent des rhetorischen Mittels der Metonymie verstanden werden. Strukturbildern liegt das Mittel der Metapher zugrunde.

3.5.5 Die nicht-gegenständlichen Bilder der modernen Kunst, die auf den ersten Blick ein Problem für die skizzierte Bildtheorie darstellen, können mithilfe des reflexiven Modus des Bildumgangs behandelt werden: Mit ihnen werden nicht einfach nur Kontexte mit Gegenständen zur Verfügung gestellt. Vielmehr sollen verschiedene Aspekte der speziellen Weise, in der Bilder funktionieren, thematisiert (und insbesondere exemplifiziert) werden. Reflexive Bilder, also Bilder, die in reflexivem Modus zu gebrauchen sind, treten aber auch außerhalb der Kunst auf, etwa in Artikeln der Computergraphik.

3.6 Eine Zusammenstellung der für Computervisualistik wichtigsten Aspekte der skizzierten Bildtheorie leiten zum Hauptkapitel über, das sich mit der informatischen Behandlung dieser Aspekte beschäftigt.

4. In diesem zentralen Kapitel wird versucht, das grundlegende Thema der Computervisualistik in einer übersichtlichen Form zu entwickeln. Es handelt sich um die generische Form der Datenstruktur, die den Datentyp »Bild« enthält. Ihre Eigenschaften werden auf alle spezifischeren Formen vererbt.

4.1 Als übersichtliches Schema, welches das Kapitel organisiert, wird die aus der Semiotik bekannte Aufgliederung in syntaktische, semantische und pragmatische Aspekte zugrundegelegt.

4.2 Syntax stellt die bei weitem eingeschränkteste Perspektive auf Zeichen zur Verfügung, da es nur um Eigenschaften von und Relationen zwischen Bildträgern geht.

4.2.1 Trotzdem hat sich in der jüngeren Vergangenheit gerade in diesem Bereich eine wichtige Diskussion der Bildwissenschaft entfaltet, die für die informatische Behandlung nicht ohne Bedeutung ist: die Frage nach der Rolle der Auflösung für die Identität von Bildern.

4.2.1.1 Mathematisch gesprochen geht es darum, ob die Syntax von Bildern dicht sein muß, was eine wichtige Auswirkung auf die Entscheidbarkeit von Vergleichen zwischen Bildern hat. Interessanterweise ist bislang die alternative Eigenschaft der Kontinuität nicht weiter diskutiert worden.

4.2.1.2 In der Informatik wurden entsprechend Ansätze zur Enkodierung von Bildsyntax unterschiedlicher Entscheidbarkeitskomplexität für verschiedene Aufgabenstellungen vorgeschlagen: von der Pixelmatrix zum generativen Bild(träger).

4.2.2 Bildliche Auflösung entspricht in etwa der Ebene der Buchstaben bei Sprache; sprachliche Syntax beschäftigt sich aber zum großen Teil mit der Zusammensetzung von Wörtern aus Silben (Morphemen), von Sätzen aus Wörtern, und von Texten aus Sätzen. Entsprechend kann auch Bildsyntax zu größeren Einheiten übergehen, den Pixemen und ihrer Komposition.

4.2.2.1 Eine solche „Bildmorphologie“ muß einerseits die „Silben“ festlegen. Pixeme sind, allgemein gesagt, mehr oder weniger zusammenhängende geometrische Einheiten. Um eine brauchbare Formalisierung zu erreichen, bietet sich eine Nicht-Standardform der Geometrie an, die Mereo-Geometrie, da sie nicht auf nulldimensionalen Punkten als Grundelementen basiert, sondern auf weniger abstrakten ausgedehnten sogenannten Individuen. Solche Individuen kann man sich zwanglos als Ergebnisse visueller Wahrnehmungsprozesse denken.

4.2.2.2 Der Vollständigkeit halber wird ein Randbereich der Bildsyntax erwähnt: die Komposition höherer Bildeinheiten aus mehreren an sich bereits vollständigen Bildern. Film und *Comics strips* komponieren auf unterschiedliche Weise in einer Dimension, Comics nutzen zwei Layout-Dimensionen, Ausstellungen und verschiedene immersive Systeme ordnen Bilder im Raum an.

4.2.3 Die geometrischen Aspekte der Syntax stellen gewissermaßen die syntaktische Basisstruktur bereit, die auf unterschiedliche Weise mit Markerwerten belegt wird, und die umgekehrt die wahrnehmungspsychologische Rekonstruktion der Pixeme basiert. Bildliche Markerwerte sind im wesentlichen die Farben. Die Formalisierung von Farbe hat historisch verschiedene Farbmodelle hervorgebracht, die sich die Informatik zunutze macht. Wieder stellt sich die Frage, inwieweit der Farbraum dicht oder gar kontinuierlich sein muß, bzw. ob diskrete Approximationen genügen. Textur wird als ein höherstufiges Markersystem identifiziert, das auf Farbe aufbaut. Transparenz und Reflexion werden selten in diesem Zusammenhang betrachtet, stellen aber ein wichtiges Moment dar, das auch die Informatik zumindest teilweise (alpha-Kanal) zur Verfügung stellt. Im Gegensatz zu der Annahme der meisten bildwissenschaftlichen Arbeiten, daß es keine

syntaktisch fehlerhaften Bilder geben könne, eröffnet die Unterscheidung zwischen geometrischer Basisstruktur und farblichen Markerwerten zudem eine Möglichkeit dafür, zu definieren, was syntaktisch fehlerhafte Bilder sind.

4.2.4 Transformationen, die nur auf syntaktischen Eigenschaften der transformierten Bilder beruhen, sind der klassische Gegenstand der Bildverarbeitung. Einige der wichtigsten Arten solcher Transformationen werden beispielhaft skizziert.

4.2.5 Syntaktische Kategorien genügen, um Bildträger im Computer zu enkodieren, sie zu speichern, zu senden, oder in den verschiedenen Ausgabemedien zu präsentieren. Sie genügen nicht, wenn der Rechner tatsächlich mit Bildern umgehen soll. Tatsächlich können selbst die höheren syntaktischen Einheiten (komplexe Pixeme) nur korrekt aus den möglichen herausgefiltert werden, wenn semantische Aspekte mit berücksichtigt werden. Die syntaktische Struktur definiert also nur einen Teil-Datentyp der generischen Gesamtstruktur.

4.3 Die semiotisch relevanten Beziehungen zwischen Zeichenträger und dem damit Bezeichneten stehen nun im Mittelpunkt: Bildbedeutung wird dabei zunächst in der Form von prädikatorischem Bildinhalt und nominatorischer Bildreferenz untersucht. Es interessieren uns insbesondere die Repräsentationsrelation (gegebene Syntax zu gesuchtem Bildinhalt) und ihr Inverses (gegebener Bildinhalt zu gesuchter Syntax). Die Problematik, dafür einen Bezug zu den Teilfunktionen von Aussagen herzustellen, wird andiskutiert.

4.3.1 Die zentrale von repräsentationalen Bildern dargestellte Kategorie sind materielle, raum-zeitliche Gegenstände, also Sortale. Das Begriffsfeld der sortalen Gegenstände bildet daher die Basis für die Komponente »Bildinhalt«. Das muß zugleich der Ausgangspunkt für die Erzeugung eines repräsentationalen Bildes sein. In der Computergraphik werden dazu allerdings geometrische Modelle verwendet.

4.3.1.1 Gemäß der Konstitutionsbeziehung zwischen Sortalen und geometrischen Individuen (Abschnitt 3.4.4) wird damit lediglich ein notwendiger Bestandteil von »Bildinhalt« angegeben, aber keineswegs der vollständige Datentyp. Der für sortale (d.h. Kontexte miteinander verbindende) Identität notwendige meronomische Anteil fehlt.

4.3.1.2 Um genauer zu verstehen, was das bedeutet, ist ein kurzer Exkurs in die Theorie der Argumentation notwendig, der sich ausdrücklich auf die argumentationstheoretische Analogie der Informatik aus Abschnitt 2.1 stützt. Rationale Argumentation wird benötigt, um einerseits das korrekte Zuschreiben eines Begriffs (d.h.: das Zutreffen einer Prädikation) zu kontrollieren, indem etwa eine Definition angewendet wird, und andererseits, um Begriffsklärungen durchzuführen. Diese äußern sich u.a. in Bestimmungen der Bedeutungspostulate, die die Verwendung der Begriffe eines Feldes beschreiben bzw. festlegen (z.B. Definitionen oder syllogistischen Schlußschemata). Die Bedeutungspostulate selbst können so nicht hinterfragt werden. Kann auf dieser Ebene kein Konsens erzielt werden, muß eine komplexere Form der Argumentation angewendet werden, bei der Beziehungen zwischen Begriffsfeldern zur Sprache kommen. Wichtig ist hier vor allem das argumentationstheoretische Analogon zur Implementierungsbeziehung: Die innere Struktur eines Begriffsfeldes wird verstanden als systematische Kombination von anderen, in der Regel einfacher strukturierten Feldern. Auf diese Weise können die Bedeutungspostulate des implementierten Feldes selbst motiviert werden – so wie analog gezeigt werden kann, daß eine bestimmte Implementierung eine Spezifikation (symbolische Beschreibung erwünschter Bedeutungspostulate) erfüllt. Außerdem können die mit den elementareren Feldern verbundenen sensomotorischen Testroutinen, mit denen Instanzen der entsprechenden Begriffe gefunden werden, auch zur referentiellen Verankerung des komplexen Feldes verwendet werden.

4.3.1.3 Übertragen auf das Feld der sortalen Objekte mit seiner überaus komplexen inneren Struktur heißt das: die oft unklaren Schlußschemata für räumliche Relationen können als das Resultat einer systematischen Kopplung von geometrischen und meronomischen Begriffen motiviert werden. Ferner können die in der Zeit ausgedehnten Instanzen sortaler Objekte mithilfe der ererbten sensomotorischen Testroutinen für die instantanen Gestalt-Individuen (unter anderem) referentiell verankert werden. Das bildet die Grundlage der referenzsemantischen Verankerung von sprachlichen Ausdrücken, die sich mit sortalen Gegenständen und ihrer relativen Lage zueinander befassen. Die geometrischen Modelle der Computergraphik beschränkt sich auf diesen Aspekt.

4.3.1.4 Die Beziehung zwischen sortalem Bildinhalt und der syntaktischen Struktur des Bildträgers, oft als „Perspektive“ bezeichnet, besteht mithin tatsächlich aus *zwei* Projektionsschritten: der Projektion des sortalen Individuums auf eine (mehr oder weniger) instantane geometrische Gegebenheitsweise in drei Dimensionen; und einer zweiten Projektion des dreidimensionalen geometrischen Modells in die beiden Dimensionen der Bildsyntax. Erstaunlichweise gehen einschlägige Arbeiten zur Perspektive nur sehr selten auf die erste, für das Verständnis des Phänomens „Bild“ offensichtlich zentrale Projektion ein. Verschiedene Abbildungsstile, wie sie mit photorealistischem bzw. nichtphotorealistischem Rendering erzeugbar sind, lassen sich durch Variationen der beiden Projektionsschritte erklären.

4.3.2 Auch das Gebiet des Bildverstehens bzw. Computersehens bedient sich der beiden Projektionsschritte, allerdings (teilweise) in umgekehrter Richtung. Grundlegend ist allerdings in der Regel zunächst, überhaupt die vollständige syntaktische Struktur nicht nur auf der Ebene der Pixel zu erzeugen. Dabei spielen die Gesetze der Gestaltgruppierung, die entsprechende Aspekte unserer visuellen Wahrnehmungsfähigkeit beschreiben, eine zentrale Rolle.

4.3.2.1 Anhand eines ausführlichen Beispiels werden wichtige Kriterien beim Aufbau von Pixemen höherer Ordnung gezeigt. Dabei wird auch deutlich, warum diese Pixeme, die in der Bildverarbeitung häufig einfach „Objekte“ genannt werden, bestenfalls mit den Präobjekten aus Abschnitt 3.3.3, nicht aber mit sortalen Objekten vergleichbar sind.

4.3.2.2 Der entscheidende Schritt besteht darin, bestimmte Pixeme höherer Ordnung als eine geometrische Instantiierung eines Objektschemas (d.h., sortalen Objektbegriffs) zu interpretieren. Als Nebenresultat können auf diese Weise z.B. auch Schatten erkannt werden. Während das Finden der Pixeme höherer Ordnung meist als datengesteuertes (*bottom-up*) Verfahren realisiert wird, handelt es sich beim Instantiieren eines Objektschemas immer um einen zielgesteuerten (*top-down*) Schritt. Das Verfahren ist allerdings empfindlich gegenüber gröberen Abweichungen von der normalen Perspektive, wie Bilder von sogenannten „unmöglichen“ Figuren demonstrieren. Der Übergang ins dreidimensionale sortale Begriffsfeld durch das Instantiieren von Objektschemata ist noch nicht der letzte Schritt, denn Bilder zeigen nicht einzelne Objekte sondern räumliche Anordnungen davon. Das Erkennen räumlicher Relationen, wie wir sie sprachlich etwa durch lokative Präpositionen angeben, hängt, wie schon in 4.3.1.3 angedeutet, ebenfalls von geometrischen und meronomischen Faktoren ab. Der Komplex der insgesamt aktivierten Begriffe des sortalen Feldes verbunden mit den zugehörigen senso-motorischen Testroutinen des konstituierenden geometrischen Feldes bildet die Instanz von »Bildinhalt« des analysierten Bildes.

4.3.2.3 Da die Darstellung der Basis bildverstehender Verfahren bislang keinen Unterschied zwischen dem Sehen des aktuellen Verhaltenskontextes und dem Sehen eines Bildes, oder auch zwischen dem Sehen und dem Modellieren des Sehens im Rechner gemacht hat, werden diese Differenzierungen nun nachgeholt. Entscheidend ist, wer hier eigentlich sieht bzw. ein Bild betrachtet.

4.3.2.4 Repräsentationale Bilder zeigen Gegenstände zwar immer als von einem bestimmten Typ sortalen Gegenstands (im Gegensatz zu einem ganz bestimmten Individuum) – etwas was von einem bestimmten Begriff kontrolliert wird bzw. was diesen Begriff exemplifiziert, so daß dieser Begriff Teil des entsprechenden »Bildinhaltes« ist; gleichwohl handelt es sich dabei stets um eine einzelne Instanz. Im Sinne der Überlegungen aus Abschnitt 3.3 ist es ein intentionales Objekt, und dieses Objekt muß der Zielpunkt der Bildreferenz-Relation sein. Um die Referenzbeziehung besser zu verstehen, wird ein kurzer Überblick zur linguistischen Referenzsemantik eingeschaltet: Hier ergibt sich die Referenzbeziehung eines (sortalen) Ausdrucks gerade über die Anbindung der sensomotorischen Testroutinen des konstituierenden geometrischen Begriffsfeldes (*vulgo*: man bekommt den Referenten, indem man im passend gewählten aktuellen Verhaltenskontext hinschaut, anfaßt etc.) Infolge der argumentationstheoretischen Einbettung der Konstitutionsbeziehung besteht nun aber der Unterschied zwischen intra-lexikalischer Semantikaufassung und referenzsemantischer Auffassung nicht darin, daß im einen Fall Sprache nicht verlassen wird, im anderen dagegen schon; sondern darin, daß Argumentationsformen unterschiedlicher Komplexität mit der jeweiligen semantischen Erläuterung zur Sprache gebracht werden sollen. Die Referenzbeziehung bei Bildern kann nicht nach demselben Muster erfolgen, denn Bilder liefern primär nur eine Gegebenheitsweise eines Gegenstandes während die sprachliche Referenz immer das Sortal insgesamt meint. Die Identität des in einem Bild gegebenen intentionalen Gegenstandes mit einem dem Betrachter anderweitig bekannten Individuum muß daher in der Regel über zusätzliche sprachliche Marker etabliert werden. Die Bildreferenz ist dem Bildinhalt untergeordnet.

4.3.3 Die Behandlung der Bildsemantik in den vorangegangenen Abschnitten verfolgte die Strategie einer sehr engen Anlehnung an linguistische Kategorien; obgleich es eine Reihe guter Gründe für diese Analogie gibt, läuft sie im Grunde den bildwissenschaftlichen Überlegungen aus Kapitel 3 zuwider, der zufolge Bildkommunikation primär kontextbildende Funktion erfüllt. Um diesen Widerspruch aufzulösen, müssen vertieft pragmatische Aspekte in die Überlegung einbezogen werden.

4.4 Pragmatik ist die umfassendste Perspektive auf Zeichen und Zeichenhandlungen, da hierbei die Beziehung zwischen der Zeichenhandlung, in der das Zeichen gebraucht wird, und den anderen damit im Zusammenhang stehenden Handlungen und Verhalten der beteiligten Zeichenhandelnden betrachtet wird. Viele Aspekte der in 4.3 vorgestellten semantischen Überlegungen sind daher wegen ihres eindeutigen Handlungsbezugs tatsächlich bereits als Teil der Bildpragmatik zu verstehen. Ein deutliches Kennzeichen für pragmatische Überlegungen ist, daß die Kommunikationssituation und die daran Beteiligten eine wichtige Rolle spielen.

4.4.1 Für die Computervisualistik liefert das Bild im Rahmen eines interaktiven Systems die charakteristischste Bildverwendungssituation. Die mithilfe von interaktiven Systemen vermittelte Kommunikation zeichnet sich durch einige Besonderheiten von anderen Medien aus. Daher werden zunächst die Charakteristika der anderen Medien kurz zusammengefaßt.

4.4.1.1 Interaktive Systeme stellen Medien der Klasse IV dar, bei denen nicht einzelne Zeichen zwischen in Raum und Zeit meist weit voneinander entfernten Zeichenhandelnden ausgetauscht werden, sondern jeweils ganze Klassen von Zeichen, von denen eines für den konkreten Fall ausgewählt wird, und zwar algorithmisch aus Handlungen des Empfängers abgeleitet, d.h. nur indirekt vom Sender selbst abhängig. Im Falle bildhafter Zeichen wird dafür der Ausdruck „Telerendering“ geprägt. Natürlichsprachliche Systeme, wie sie im Gebiet „Künstliche Intelligenz“ der Informatik entwickelt werden, gehören ebenfalls zu den Medien der Klasse IV und bieten einige Anhaltspunkte für den besonderen Umgang mit pragmatischen Aspekten, der hierbei notwendig wird. Allerdings

spielt die Funktion „Kontextbilder“ dort gegenüber Prädikation und Nomination eine sehr untergeordnete Rolle. Diese Auffassung ist auch bei den meisten der derzeit bestehenden Systemen zum Telerendering am Werk, die in den folgenden Abschnitten skizziert werden.

4.4.1.2 Die Unterscheidung zwischen „what to say“ und „how to say“ läßt sich für Bilder analog zur Sprachgenerierung stellen. Auswahl des Inhalts wird analog als Bestimmung einer Instanz von »Bildinhalt« verstanden, also als eine Menge von sortalen Begriffen (Gegenstände und räumliche Relationen), die mit den zugehörigen geometrischen Projektionen assoziiert sind – i.w. den geometrischen Objektmodellen. Sollen ganz bestimmte individuelle Objekte, die aus anderem Zusammenhang bekannt sind, im Bild auftreten, muß die Koreferenz in der Regel über zusätzliche meist sprachliche Zeichen (Bildtitel u.ä.) vermittelt werden.

4.4.1.3 Auswahl der Form betrifft vor allem die Wahl des Betrachterstandpunkts und Bildausschnitts, aber auch die Beleuchtung und den Darstellungsstil. Vor allem mit den letzten beiden Parametern lassen sich „rhetorisch angereicherte“ Bilder herstellen: Obwohl Bildern als Kontextbildern an sich noch keine Aufgliederung in nominative und prädikative Teile zukommen, legen gewisse Beleuchtungen oder Darstellungsstile eine bestimmte Zuordnung nahe – die reine Funktion des Kontextbilders wird also rhetorisch überformt.

4.4.1.4 Sofern kein Bild neu generiert wird, sondern ein möglichst gut passendes aus einer gegebenen Kollektion gewählt werden soll, treten Inhalts- und Formfestlegung in einer kombinierten Fassung auf. Anhand eines Beispiels wird gezeigt, wie hierfür ein rein propositional geprägtes Semantikverständnis genutzt werden kann.

4.4.2 Eine zweite Analogie zu den Verfahren der natürlichsprachlichen Systeme ist das Berücksichtigen der Wirkung auf den oder die Zeichenhandlungspartner. Formal ergibt sie sich aus den Überlegungen von MEAD zu komplexem Zeichengebrauch, referiert in Abschnitt 3.5.3, und den sprachphilosophischen Überlegungen zu Sprechakten. Ausgangspunkt dafür ist eine Angabe des Zwecks, der mit einem Bildzeigeakt verfolgt wird.

4.4.2.1 Sprechhandlungen, und ihre verallgemeinerte Form: Zeichenakte, bilden Systeme zum regelgeleiteten Ableiten von Sequenzen von Handlungen. Rhetorische Relationen zwischen Zeichenakten beschreiben die Funktion einer Zeichenhandlung in der Sequenz. Damit eng verbunden sind explizite Zuschreibungen von Zwecken zu den Zeichenakten. Bisher wurden Bildhandlungen im wesentlichen als Substitute für assertionale Sprechakte betrachtet: sie treten mit den gleichen illokutionären Rollen auf; die rhetorischen Funktionen sind weitgehend dieselben. Mithilfe von strategischen Regeln lassen sich daher auch Planungssysteme für multimedial gemischte Zeichen formulieren, die aus vorgegebenen Zwecken rhetorische Relationen und schließlich koordinierte Bildzeige- und Sprechakte ableiten.

4.4.2.2 Der intendierte Zwecke muß keineswegs auch korrekt beim Gegenüber ankommen. Zu diesem Zweck werden in natürlichsprachlichen Systemen explizite oder implizite Partnermodelle verwendet, die für das wechselseitige Einhalten der für das Sprachspiel gültigen Regeln sorgen, etwa der Konversationsmaxime nach GRICE. Insbesondere werden beim Gegenüber durch eine Äußerung ausgelöste Implikaturen und Präsuppositionen berücksichtigt. Das funktioniert auch, wenngleich mit einigen zusätzlichen Komplikationen, für Telerendering, so daß das interaktive System die generierten Bilder etwa dem Kenntnisstand und den Erwartungen des Benutzers anpassen kann. Es wird unterschieden zwischen dem „Modell des aktiven Betrachters“, das ein Betrachter vom Sender der bildlichen Botschaft hat, und dem „Modell des passiven Betrachters“, das sich der Sender von den potentiellen Rezipienten macht. Explizite Partnermodellierung ist sehr aufwendig und wird häufig durch implizite Regeln approximiert.

4.4.2.3 Eine direkte Übertragung der linguistischen Partnermodellierung auf die Bildkommunikation in Medien der Klasse IV geht von einer Analogie zwischen Propositionen und Bildsemantik aus, nicht von der grundlegenderen Kontextbilder-Funktion. Eine Anpassung der Benutzermodellierung für Kontextbilder erfordert es, nicht von einem einmal festgelegten festen Inhalt (und assoziierter Referenz) auszugehen, sondern diese Interpretationen stets offen und revidierbar zu halten. Hierbei ist die Analyse des der Bildverwendung zugrundeliegenden immersiven Reflexionsmodus als Kombination von spontanem dezeptivem Modus mit dem symbolischen Modus aus Abschnitt 3.5.1 hilfreich. Sie verweist zugleich darauf, daß die rein semantische („unpragmatische") Auffassung von »Bildinhalt« unvollständig ist, da der symbolische Modus überhaupt nur in der Koordination zwischen aktiven und passivem Betrachtermodell sinnvoll ist. Betrachtermodelle sind nicht einfach zusätzliche Teile, sondern essentieller Bestandteil der Datenstruktur.

4.4.3 Eine zentrale Frage im Rahmen der Bildpragmatik ist die nach der Authentizität von in Medien der Klasse IV verwendeten Bildhandlungen.

4.4.3.1 Authentizität im allgemeinen kann von den Betrachtermodellen nicht gewährleistet werden, da eine direkte Beobachtung der Beteiligten nicht möglich ist. Dies gilt insbesondere für den Konstrukteur, der als primärer Sender betrachtet wird. Lediglich für den Endnutzer, der sich selbst Bilder mit einem interaktivem System zeigt und daher als sekundärer Sender betrachtete werden kann, können Interaktionsparameter dazu verwendet werden, das entsprechende Betrachtermodell zu modifizieren.

4.4.3.2 In einem wesentlich engeren technischen Sinn wird der Ausdruck „Authentizität" in der Informatik dazu benutzt, syntaktische Methoden der Übertragungssicherung zu beurteilen. Hierbei spielen insbesondere digitale Wasserzeichen eine wichtige Rolle.

4.4.4 Strukturbilder, die durch eine metaphorische Bedeutungsverschiebung erklärt werden, treten in der Computervisualistik vor allem im Bereich Informationsvisualisierung auf. Häufig helfen sie dabei, überhaupt erst eine relevante Konzeptualisierung von Rohdaten zu finden. Diese explorative Bildbenutzung beim „*data mining*" entspricht eher einem Selbstgespräch als der normalen zweiseitigen Kommunikation; sie deckt sich zugleich besser mit der Funktion als Kontextbilder, weniger mit der als Proposition (oder einer Teilfunktion davon).

4.4.4.1 Die metaphorische Bedeutungsverschiebung für Strukturbilder unterliegt bestimmten Restriktionen, die sich aus der Struktur von »Bildträger« und »Bildinhalt« ergeben: die Quelldomäne der Metapher entspricht direkt dem geometrischen Feld oder indirekt dem sortalen Feld. Das Problem von Koreferenzen in verschiedenen Bildern einer Serie stellt sich für die beiden Fälle jeweils unterschiedlich. Insofern die Zieldomäne der Metapher Subdomänen aus dem sortalen oder geometrischen Bereich (oder dazu isomorphen Begriffen) enthalten kann, kann man Strukturbilder auch einteilen in solche, die eine natürliche (repräsentationale) Kernabbildung beinhalten, und andere, für die das nicht gilt. Aus der Kombination zwischen den beiden Ausprägungen der Quelldomäne und den unterschiedlichen Zieldomänklassen lassen sich insbesondere auch repräsentationale und nonfigurative Bilder als Sonderfälle bestimmen. Auf jeden Fall müssen die nicht durch triviale Isomorphien abgedeckten Begriffe der Zieldomäne auf geeignete, d.h. für den intendierten Betrachter nachvollziehbare, konventionelle Weise auf Begriffe der Quelldomäne abgebildet, d.h., auf einen sortalen »Bildinhalt« oder direkt die geometrische Bildsyntax projiziert werden.

4.4.4.2 Als zentrale semantische Einschränkung für die Wahl einer konventionellen Zuordnung gilt, daß Quell- und Zielbegriff von gleicher „Mannigfaltigkeit" sind und dieselbe Art von Ordnungsrelationen gelten. Mißachtung dieses Grundsatzes führt leicht zu Abbildungsartefakten oder unvollständigen Projektionen. Beides kann beim Betrachter feh-

lerhaftes Verständnis auslösen. Problematisch wird die konventionelle Zuordnung allerdings vor allem dadurch, daß die Quelldomäne nur eine recht beschränkte Auswahl an möglichen Quellattribute verschiedener Ausprägungen zur Verfügung stellt. Zudem müssen häufig nicht nur viele verschiedene Zielattribute, sondern auch große Mengen an einzelnen Ausprägungen betrachtet werden. Eine Zusammenfassung mehrerer Attribute in Ikonen oder Glyphen bzw. eine texturartige Verschmelzung sehr dicht gesetzter Glyphen kann den Problemen zu einem gewissen Grad abhelfen. Auch eine zeitliche Entwicklung zu betrachten vergrößert den Spielraum, wobei allerdings eine feste Zuordnung eines Parameters zur Zeitachse die mit der Animation gegebenen Möglichkeiten nicht voll ausnutzt.

4.4.4.3 Wesentlich günstiger ist eine interaktiv wechselnde Zuordnung von Zeitachse und Zielattribut. Um große Datenmengen in interaktiver Form überschaubar zu präsentieren, sind zudem Fokus/Kontext-Mechanismen wichtig, die wiederum nur in enger Beziehung zum Vorwissen und den Intentionen der Betrachter sinnvoll zu steuern sind.

4.4.5 Reflexive Bilder spielen im Rahmen der Informatik auf zweierlei Weise eine Rolle: einerseits treten sie als Bildzitate auf, mit denen weniger der primäre Bildinhalt vorgeführt wird. Vielmehr wird das zur Erzeugung verwendete bild-generierende oder bildmanipulierende Verfahren exemplifiziert. Andererseits sind künstlerische Bilder in aller Regel reflexiv, so daß die Produkte der Computerkunst ebenfalls hier betrachtet werden müssen.

4.4.5.1 Im Unterschied zu den beiden anderen Grundkategorien, den repräsentationalen Bildern und den Strukturbildern, die sich voneinander in der Beziehung zwischen Syntax und Semantik unterscheiden, zeichnen sich reflexive Bilder durch den besonderen Rezeptionsmodus des Betrachters aus. Folglich kann jedes Bild zum reflexiven Bild werden, wenn es entsprechend rezipiert wird. Eine bestimmte Gruppe von Strukturbildern sticht in diesem Zusammenhang hervor: da bei ihr die semantische Relation zu einer Identitätsrelation entartet ist, erscheinen diese Bilder oft „unsemantisch“ (bedeutungslos), wodurch leicht eine Reflexion auf die mit ihnen potentiell zu vollziehenden Bildkommunikationshandlungen ausgelöst wird. Sie lassen sich also *nur* im reflexiven Modus verstehen.

4.4.5.2 Eben solche Bilder standen auch am Anfang der Computerkunst. Der Ausdruck ist keine Stilbezeichnung. Mit ihm werden alle Werke der bildenden Kunst zusammengefaßt, bei der ein Künstler mit dem Computer als Werkzeug gearbeitet hat. Die beim Betrachten des Werks ausgelöste Reflexion auf Bildkommunikation reicht beispielsweise von einer algorithmischen Untersuchung des Zusammenhangs zwischen den vielen Perspektiven eines Sortals und der einen davon in einem konkreten Kontext verfügbaren geometrischen Projektion (ähnlich dem Kubismus) bis zu Analyse kognitiver Fähigkeiten, die beim Umgehen mit Bildern unumgänglich sind, mit den Mitteln der KI.

4.4.5.3 Auch im Zusammenhang der Computerkunst offenbart sich der eigentliche Beitrag des Werkzeugs Computer in der Interaktivität. Diese forciert geradezu eine deutliche Verschiebung des in Europa tradierten Werkbegriffs: Ein bildliches Kunstwerk ist nicht notwendig an einen individuellen materiellen Träger gebunden. Mit dem Aufstieg der Computergraphik sind insbesondere immersive Kunstwerke möglich geworden, die auf den ersten Blick zur antiken Kunstauffassung bzw. dem Ideal eines absoluten *trompe l'œil* zurückzukehren scheinen: der Zeichencharakter der Bilder scheint völlig zu verschwinden. Allerdings stellen die gelungenen Kunstwerke dieser Art die Täuschung einiger Nutzer stets in einen explizit reflexiven Zusammenhang für viele andere Betrachter. Der gespannte Zusammenhang zwischen der Wahrnehmungsnähe und dem Zeichencharakter der Bilder wird letzteren ausdrücklich vorgeführt. Ausnahmen davon können in der Regel als Kunstwerke höhere Ordnung betrachtet werden, bei denen die

eingesetzten Bilder selbst gar nicht im Vordergrund stehen, sondern in ein Gesamtkunstwerk aus vielen anderen Komponenten einfließen. Ihre Verwendung ist entsprechend auch gar nicht Ziel der künstlerischen Reflexion.

5. Vier Fallbeispiele ergänzen die Überlegungen des vorigen Kapitels

5.1 Nur syntaktische und semantische Aspekte der generischen Datenstruktur kommen zum Einsatz, wenn es darum geht, nach Bildern inhaltlich in einer Bilddatenbank suchen zu können.

5.1.1 Das vorgestellte System IRIS – Image Retrieval for Information Systems – exemplifiziert den Einsatz bildanalytischer Algorithmen und eines stark vereinfachten Schemas der Objektkonstitution.

5.1.2 Ein Beispiel der Analyse und die Möglichkeiten der Suchanfragen werden gezeigt

5.2 Werden syntaktische Aspekte verwendet, um pragmatische Faktoren zu variieren, so sprechen wir von rhetorisch angereicherten Bildern.

5.2.1 Obwohl die Interpretation eines wahrnehmungsnahen Kontextbilders nicht absolut festliegt, können durch geschickt gewählte stilistische Variationen im Bild bestimmte Interpretationen in einem rhetorischen Sinn betont werden.

5.2.2 Besonders grundlegend ist dabei, gezielt eine eindeutige Zuordnung des Bildes zu den beiden grundlegenden kommunikativen Teilfunktionen der Aussage, Nomination und Prädikation, herbeiführen zu können.

5.2.3 Zu diesem Zweck wird die Heuristik des prädikativen Naturalismus vorgeschlagen und an einem Beispiel demonstriert.

5.3 Einen Sonderfall von Bildern stellen die immersiven Bilder dar, da bei ihnen – zumindest bei einem Teil der Nutzer – der Zeichenaspekt unterdrückt wird. Allerdings wird bei genauerer Betrachtung sinnvoller Anwendungen sehr schnell klar, daß beide Bildaspekte eine wichtige Rolle spielen, deren Gewicht sich je nach Teilaufgabe verändert.

5.3.1 Ein Beispiel für den sinnvollen Einsatz immersiver Systeme ist die virtuelle Architektur: Beispielhaft wird das Atmosphären-Projekt vorgestellt, in dem das teilweise verlorene Haus Behrens mit digitalen Mitteln rekonstruiert wird. Bei diesem einzigartigen Jugendstilensemble, das der spätere Begründer des Industriedesigns, PETER BEHRENS, 1901 zu einer Architekturausstellung in Darmstadt als sein Wohnhaus entworfen hat, kommt ein Ideal des Jugendstils besonders deutlich zur Ausprägung: der Versuch, alle Lebensvollzüge durch eine æsthetische Gestaltung der Umwelt positiv zu verändern. Da die computervisualistische Umsetzung speziell hinsichtlich dieser Zielsetzung untersucht werden soll, spielt das harmonische Zusammenwirken aller Elemente auf die Atmosphäre der Räumlichkeiten eine wichtige Rolle. Störende Artefakte, die aus der Modellierung oder unklarer Quellenlage resultieren, müssen unbedingt vermieden werden. Zugleich muß deutlich bleiben, welche Texturen etwa dem verlorenen Original wirklich genau entsprechen, und welche anderen nur ein mehr oder weniger plausibler Ersatz sind.

5.3.2 Allerdings kann die Rekonstruktion in ganz verschiedenem Zusammenhang eingesetzt werden. Insbesondere müssen bildende, wissenschaftliche und unterhaltende Verwendungskontexte unterschieden werden. Jede davon stellt unterschiedliche Anforderungen daran, inwieweit eine ungestörte Immersion für den Benutzer aufgebaut werden muß, oder zu welche Grad der Zeichencharakter der Bilder überhaupt erst wichtige Handlungsgrundlagen abgibt.

5.3.3 Obzwar von der Aufgabenstellung her ganz verschieden, zeigt sich auch für das zweite Beispiel, das virtuelle Institut für Bildwissenschaft, daß die erforderlichen Funktionen nur dann sinnvoll gewährleistet werden können, wenn zu bestimmten Zwecken ein hoher Grad an Immersion erreicht wird, der für andere Zwecke geradezu störend wirkt.

5.3.4 Zusammengefaßt hängt also der Einsatz von Bildern auch in immersiven Systemen von der Doppelnatur perzeptoider Zeichen ab: in einigen Fällen, wie dem der virtuellen Architektur, fluktuiert das Gewicht der beiden Aspekte Täuschung und symbolische Distanz durch den Betrachter im Rahmen des Gebrauchs. In anderen Fällen, wie dem der virtuellen Institute, sind immersive Bilder nur für bestimmte Teilaufgaben sinnvoll.

5.4 Eine noch peripherere Verwendung des Ausdrucks „Bild“ liegt bei mentalen Bildern vor. Ob die Bildwissenschaft hier überhaupt noch zuständig ist, bzw. ob die Computervisualistik helfen kann, ist auf den ersten Blick eher zweifelhaft**.**

5.4.1 Ein typisches Beispiel für die Rede von mentalen Bildern bieten referenzsemantische Erklärungen des Verstehens von Berichten über weit entfernte räumliche Gegebenheiten.

5.4.2 Ein kurzer Exkurs über die kognitive Funktion von mentalen Bildern eröffnet trotz fundamentaler Unterschiede eine große Ähnlichkeit in der Struktur des Redens über Bilder bzw. über mentale Bilder, d.h., zwischen den beiden Begriffen.

5.4.3 Als Beispiel wird die Architektur eines Systems kurz skizziert, das die Argumente des Erklärungsansatzes des Verstehens räumlicher Berichte mittels Bildvorstellungen exemplifiziert. Dieser Entwurf wurde im System ANTLIMA praktisch umgesetzt, um ein referenzsemantisches Hörermodell zu implementieren.

5.4.4 Insgesamt gesehen können wesentliche Elemente der generischen Datenstruktur »Bild« deshalb auch in Aufgabenstellungen zu mentalen Bildern Anwendung finden, weil sich die Informatik vor allem mit Begriffen beschäftigt, nämlich in der Gestalt abstrakter Datentypen. Mindestens die begrifflichen Zusammenhänge, die syntaktischen und semantischen Aspekte von »Bild« betreffen, strukturieren auch die Rede von mentalen Bildern.

6. Zusammenfassung und Perspektiven

6.1 Kehren wir zuletzt zu der Argumentation aus Kapitel 1 und dem Ziel der ganzen Untersuchung zurück: Der Zweck einer genauen und umfassenden Analyse der wesentlichen Bestandteile der Datenstruktur um den Datentyp »Bild« und ihrer Anwendungsbedingungen bestand darin, den Kern der sich gerade herausbildenden Computervisualistik, der Bildwissenschaft der Informatik, bereitzustellen.

6.2 Ein wichtiges Element hin zu einer institutionalisierten Computervisualistik besteht darin, eine entsprechende Ausbildung anzubieten.

6.2.1 Hierin nimmt die Magdeburger Universität die Führungsrolle ein. Im Licht der vorangegangenen Untersuchung wird kurz die Struktur des Studiengangs Computervisualistik dargestellt.

6.2.2 In diesem Zusammenhang stellt sich die Frage, ob bzw. inwiefern sich der Studienerfolg durch einen Zugangstest abschätzen läßt. Besonders die Fähigkeit zur Raumvorstellung wird hierbei von Psychologen ins Auge gefaßt, ein Parameter, der allerdings in hohem Maße geschlechtsabhängig ist.

6.2.3 Aus diesem Grund werden die Ergebnisse einer ersten empirischen Studie vorgestellt, bei der Computervisualistik-Studierende als eine der wenigen berichteten Gruppen ohne signifikante Geschlechtsunterschiede bei der Raumvorstellung hervortraten.

6.3 Die vorliegende Arbeit kann nicht mehr sein als eine grobe Landkarte mit wenigen, auffälligen Landmarken, die der zukünftigen Forschung im Rahmen einer einheitlichen Computervisualistik im Überlappungsbereich von Informatik und allgemeiner Bildwissenschaft Orientierungshilfen geben soll. Es wird der Hoffnung Ausdruck gegeben, daß die Arbeit in diesem Sinne erfolgreich wirken möge.

About the Author

Jörg R.J. Schirra was born in Saarland, Germany, in 1960. He studied computer science, physics, psychology, and philosophy at the University of Saarland at Saarbrücken, where he also worked, after having received his diploma in computer science, as a researcher in the Special Collaborative Project "Artificial Intelligence and Knowledge-Based Systems" of the German Research Council (DFG) until 1993. In 1994, he received his doctoral degree in computer science there. Subsequently, he spent one year as a post-doctoral fellow at the International Computer Science Institute (ICSI) at Berkeley, California. At the University of Magdeburg, where he worked from1996 till 2003, he was responsible for the diploma programme in Computational Visualistics.

Jörg R.J. Schirra

3. August 1960	geboren in Illingen, Saarland, BRD
Juni ’79	Abitur am Pädagogium Baden-Baden
Oktober ’79 — Oktober ’86	Studium der Informatik mit Nebenfach Physik an der Univ. d. Saarlandes, Saarbrücken
	in dieser Zeit Besuch zusätzlicher Kurse in Kognitionspsychologie (Prof. Tack); Psycholinguistik (Prof. Engelkamp), Philosophie (Prof. Kuno Lorenz und P.D. Arno Ros), philosophischer Logik (Prof. Risse) und chinesischer Sprache (Dr. Xia)
Oktober ’86	Abschluß als Diplom-Informatiker; Diplomarbeit bei Prof. W. Wahlster über das Thema *MEGA-ACT — Eine Studie über explizite Kontrollstrukturen und Pseudoparallelität in einem Aktor-System mit einer Beispielarchitektur in FRL*
Oktober ’86 — November ’93	wissenschaftlicher Mitarbeiter im Projekt VITRA (Visual Translator) im SFB 314 „KI und Wissensbasierte Systeme“ an der Univ. d. Saarlandes
	Koordination des interdisziplinären PHILIP-Kreises zur Kognitionswissenschaft an der Univ. d. Saarlandes
März ’94 — Dezember ’95	wissenschaftlicher Mitarbeiter der Arbeitsgruppe „Künstliche Intelligenz“ im Fachbereich Informatik der Universität Bremen im Drittmittelprojekt *Image Retrieval in Information Systems (IRIS)*; unterbrochen durch den Aufenthalt in Berkeley (s.u.)
Juli ’94	Dissertation zum „Doktor rerum naturalium“ der Technischen Fakultät der Universität des Saarlandes bei Prof. W. Wahlster über das Thema *Bildbeschreibung als Verbindung von visuellem und sprachlichem Raum — Eine interdisziplinäre Untersuchung von Bildvorstellungen in einem Hörermodell.*
September ’94 — September ’95	Postdoc-Stipendium am *International Computer Science Institute*, Berkeley, California, USA
Januar ’96 — April 2003	wissenschaftlicher Mitarbeiter am *Institut für Simulation und Graphik* der *Fakultät für Informatik* an der *Otto-von-Guericke-Universität Magdeburg*
	seit ’97 wissenschaftlicher Assistent (zur besonderen Verwendung)
	hier u.a. Koordination des interdisziplinären Bildwissenschaftlichen Kolloquiums;
	verantwortlich für Einführung, Organisation und Koordination des Diplomstudiengangs Computervisualistik; zuständig für interfakultäre Abstimmung sowie *public relations*; zudem Studienfachbetreuer für Computervisualistik
Seit April 2003	Unabhängiger Privatgelehrter
Oktober 2003	Fertigstellung der Habilitationsschrift, angenommen im November 2003 von der Fakultät für Informatik der OvG-Universität Magdeburg
Februar 2005	Eingang des letzten Gutachtens zur Habilitationsschrift
11. Mai 2005	Erfolgreiche Verteidigung, Abschluß des Habilitationsverfahrens

Wissenschaftliche Publikationen bis 2005

A) Monographie & Buchkapitel

1. Computervisualistik: Die Bildwissenschaft der Informatik. In: K. Sachs-Hombach (Ed.): *Bildwissenschaft. Disziplinen, Themen, Methoden.* Frankfurt/M.: Suhrkamp, 2005 (ISBN: 3-518-29351-6), 268–280.
2. Von Bildern und neuen Ingenieuren: Computervisualistik als Studienfach. (JRJS[*], Klaus Sachs-Hombach) In: U. Reinhard, U. Schmid (Hrsg.): *Who Is Who in Multimedia Bildung '98*. Heidelberg: Whois-Verlag, 1998 (ISBN 3-980-4968-4-8), 226–231.
3. Abstraction versus Realism: Not the Real Question. (JRJS, Martin Scholz) Kapitel 22 in Strothotte et al.: *Computer Visualization — Graphics, Abstraction, and Interactivity.* Heidelberg: Springer, 1998 (ISBN: 3-540-63737-0), 379–401.
4. Von Bildern und neuen Ingenieuren: Aspekte des Studiengangs Computervisualistik. (JRJS, Thomas Strothotte) In: A. Dress, G. Jäger (eds.) *Visualisierung in Mathematik, Technik und Kunst: Grundlagen und Anwendungen*. Braunschweig/Wiesbaden: Vieweg, 1998 (ISBN: 3-528-06912-0), 189–205.
5. *Bildbeschreibung als Verbindung von visuellem und sprachlichem Raum — Eine interdisziplinäre Untersuchung von Bildvorstellungen in einem Hörermodell.* St. Augustin: Infix, 1994 (ISBN 3-929037-71-8), xii+446, 195 Abbildungen.

B) Begutachtete Zeitschriftenartikel

1. Ein Disziplinen-Mandala für die Bildwissenschaft - Kleine Provokation zu einem Neuen Fach. IMAGE: Journal of Interdisciplinary Image Science. 2005, 1(1), (ISSN: 1614-0885, www.image-online.info)
2. Von der interdisziplinären Grundlagenforschung zur computervisualistischen Anwendung. Die Magdeburger Bemühungen um eine allgemeine Wissenschaft vom Bild. (K. Sachs-Hombach, JRJS) In: Magdeburger Wissenschaftsjournal, 2002(1): 27–38, (ISSN: 0949-5304).
3. Motivationsanreize zum Erwerb kommunikativer Kompetenzen: Eine Fallstudie in der Informatik. In: Global Journal of Engineering Education, Special Edition "German Network of Engineering Education II", 5(3): 311–318, 2002 (ISSN: 1328-3154).
4. Sind Studentinnen der Computervisualistik besonders gut in der Raumvorstellung? Psychologische Aspekte bei der Wahl des Studienfachs. (C. Quaiser-Pohl, W. Lehmann, JRJS) In: Fiff-Kommunikation 18(3):42–46, 2001 (ISSN: 0938-3476).
5. A New Theme for Educating New Engineers: Computational Visualistics. In: *Global Journal of Engineering Education*, 4(1):73–82, 2000. (ISSN 1328-3154)
6. Von der Bildwissenschaft zur Computervisualistik. (K. Sachs-Hombach, JRJS) In: *Semiotische Berichte* (mit: *Linguistik Interdisziplinär*) – *Ikonische Zeichen: Auswahlakten zweier Konferenzen*, 24(1-4):287–304, 2000 (ISSN: 0254-9271).
7. Referenzsemantik in der Partnermodellierung. In: *Kognitionswissenschaft,* 6(4):177–195, 1997. (ISSN 0938-7986)
8. Deklarative Programme in einem Aktor-System: MEGA-ACT. In: *KI — Künstliche Intelligenz: Forschung, Entwicklung, Erfahrungen,* 3:4–9 und 4:4–12, 1988. (ISSN 0933-1875)
9. From Image Sequences to Natural Language: A First Step towards Automatic Perception and Description of Motions. (JRJS, G. Bosch, C.K. Sung, G. Zimmermann) In: *Applied AI,*1(3):287–305, 1987. (ISSN 0883-9514)

[*] Zur Vereinfachung wurde im folgenden bei Ko-Autorschaft das Kürzel „JRJS" für „Jörg R.J. Schirra" verwendet (siehe auch *http://www.JRJS.de/Work/Papers/)*.

C) ANDERE BEGUTACHTETE SCHRIFTEN

1. Begriffsgenetische Betrachtungen in der Bildwissenschaft: Fünf Thesen. In: K. Sachs-Hombach (Hrsg.): Bild und Medium. Kunstgeschichtliche und philosophische Grundlagen der interdisziplinären Bildwissenschaft. Frankfurt/M.: Suhrkamp, 2005 (in press).
2. Computational Visualistics: Dealing with Pictures in Computer Science. In: K. Sachs-Hombach (ed.): *Bildwissenschaft zwischen Reflexion und Anwendung*. (ISBN: 3-931606-73-2) Köln: Herbert von Halem Verlag, 2005, 494–509.
3. Virtual Institutes: Between Immersion and Communication. (K. Sachs-Hombach, JRJS, J. Schneider) In: W. Marotzki, J. Schneider, Th. Strothotte (eds.): Proc. *Computational Visualistics, Media Informatics, and Virtual Communities.* 2003 (ISBN 3-8244-4550-6). Deutscher Universitätsverlag, 37–50.
4. Bildwissenschaft und Computervisualistik. (K. Sachs-Hombach, JRJS) In*: Proceedings des 7. Internationalen Kongresses der IASS* (in press).
5. Selecting Styles for Tele-Rendering: Toward a Rhetoric in Computational Visualistics. (K. Sachs-Hombach, JRJS) In: *Proc. of the 2nd International Symposium on Smart Graphics* 2002, Hawthorne N.Y.: ACM Press, 102–106.
6. Identifikationsformen in Computerspiel und Spielfilm. (JRJS, St. Carl-McGrath) In: M. Strübel (Hrsg.): *Film und Krieg - Die Inszenierung von Politik zwischen Apologetik und Apokalypse*, Leverkusen: Leske+Budrich, 2002 (ISBN: 3-8100-3288-3), 149–163.
7. Content-Based Reckoning for Internet Games. In: *GAME-ON 2001 – Proc. of the 2nd International Conference on Intelligent Games and Simulation.* Ghent: SCS, 2001 (ISBN: 90-77039-04-X), 107–111.
8. Bilder ⟶ Kontextbilder. In: K. Sachs-Hombach (Hrsg.): *Bildhandeln – Interdisziplinäre Forschungen zur Pragmatik bildhafter Darstellungsformen*, Magdeburg: Skriptum, 2001 (ISBN: 3-933046-38-6), 77–100.
9. Das Haus als Gesamtkunstwerk -- eine Herausforderung an die Computervisualistik. (K. Buchholz, JRJS) In: K. Sachs-Hombach (Hrsg.): *Bildhandeln – Interdisziplinäre Forschungen zur Pragmatik bildhafter Darstellungsformen*, Magdeburg: Skriptum, 2001 (ISBN: 3-933046-38-6), 241–268.
10. Computervisualistik als angewandte Bildwissenschaft. (K. Sachs-Hombach, JRJS) In: Hess-Lüttich, Ernest W.B. (Hrsg.): *Medien, Texte und Maschinen - Angewandte Mediensemiotik.* Wiesbaden: Westdeutscher Verlag, 2001 (ISBN: 3-531-13622-4), 117–137.
11. "Computer game design": How to motivate engineering students to integrate technology with reflection. In: Z.J. Pudlowski (Hrsg.): *Proc. 4th Annual Conference of the UNESCO International Centre for Engineering Education*, Bangkok, 2001 (ISBN: 0 7326 2146 1), 165–169.
12. Täuschung, Ähnlichkeit und Immersion — Die Vögel des Zeuxis. In: K. Sachs-Hombach, K. Rehkämper (Hrsg.): *Vom Realismus der Bilder: Interdisziplinäre Forschungen zur Semantik bildhafter Darstellungsformen*, Magdeburg: Scriptum, 2000, 119–135.
13. Computational Visualistics: Bridging the Two Cultures in a Multimedia Degree Programme. Keynote address at the *2nd Asia-Pacific Forum on Engineering & Technology Education*, Sydney, 4. – 7. Juli, 1999. In: Z. J. Pudlowski (Hrsg.): Forum Proceedings, Monash Univ. Press (ISBN 0-7326-1784-7), 47–51. *Das Papier wurde mit dem* UICEE Silver Award for an outstanding paper *ausgezeichnet.*
14. Zur politischen Instrumentalisierbarkeit bildhafter Repräsentationen. Philosophische und psychologische Aspekte der Bildkommunikation. (Klaus Sachs-Hombach, JRJS) In: Wilhelm Hofmann (Hrsg.): *Die Sichtbarkeit der Macht — Theoretische und empirische Untersuchungen zur visuellen Politik*. Baden-Baden: Nomos, 1999 (ISBN 3-7890-6292-8), 28–39.
15. Syntaktische Dichte oder Kontinuität — Ein mathematischer Aspekt der Visualistik. In: K. Sachs-Hombach, K. Rehkämper (Hrsg.): *Bildgrammatik. Interdisziplinäre Forschungen zur Syntax bildlicher Darstellungsformen*, Magdeburg: Scriptum, 1999 (ISBN 3-933046-24-6), 103–119.

16. Zwei Skizzen zum Begriff 'Photorealismus' in der Computergraphik. (JRJS, Martin Scholz) In: K. Sachs-Hombach, K. Rehkämper (Hrsg.): *Bild - Bildwahrnehmung - Bildverarbeitung: Interdisziplinäre Beiträge zur Bildwissenschaft*. Wiesbaden: Deutscher Universitätsverlag, 1998 (ISBN: 3-8244-4303-1), 69–79.
17. Computervisualistik: Ein Diskussionsbeitrag zur universitären Ausbildung im Bereich Multimedia. (Thomas Strothotte, JRJS) In: J. Dassow, R. Kruse (Hrsg.): *Informatik '98* – Informatik zwischen Bild und Sprache., Proc. der 28. Jahrestagung der Gesellschaft für Informatik 1998, Berlin: Springer, 365–376.
18. Optional Deep Case Filling and Focus Control With Mental Images: ANTLIMA-KOREF. (Anselm Blocher, JRJS) In: *Proc. of IJCAI-95*, Montreal, 1995, 417–423.
19. Understanding Radio Broadcasts On Soccer: The Concept `Mental Image' and Its Use in Spatial Reasoning. In: K. Sachs-Hombach (Hrsg.): *Bilder im Geiste: Zur kognitiven und erkenntnistheoretischen Funktion piktorialer Repräsentationen*. Amsterdam: Rodopi, 1995 (ISBN:90-5183-679-1), 107–136.
20. ANTLIMA — A Listener Model with Mental Images. (JRJS, Eva Stopp) In: *Proceedings of IJCAI-93,* Chambéry, 1993, 175–180,
21. Connecting Visual and Verbal Space: Preliminary Considerations Concerning the Concept 'Mental Image'. In: M. Aurnague, A. Borillo, M. Borillo, M. Bras (Hrsg.): *Semantics of Time, Space, and Movement.* Proceedings, Toulouse: University Paul Sabatier, IRIT, 1993, 105–121.
22. A Contribution to Reference Semantics of Spatial Prepositions: The Visualization Problem and its Solution in VITRA. In: Cornelia Zelinsky-Wibbelt (Hrsg.): *The Semantics of Prepositions — From Mental Processing to Natural Language Processing*. Berlin: Mouton de Gruyter, 1993 (ISBN 3-11-013634-1), 471–515.
23. Expansion von Ereignis-Propositionen zur Visualisierung — Die Grundlagen der begrifflichen Analyse von ANTLIMA. In: H. Marburger (Hrsg): *GWAI-90 — German Workshop on AI,* Berlin: Springer, 1990, 246–256.
24. Einige Überlegungen zu Bildvorstellungen in kognitiven Systemen. In: C. Freksa, C. Habel (Hrsg.): *Repräsentation und Verarbeitung räumlichen Wissens,* Berlin: Springer, 1990*,* 68–82.
25. Ein erster Blick auf ANTLIMA — Visualisierung statischer räumlicher Relationen. In: D. Metzing (Hrsg.): *GWAI-89 — German Workshop on AI,* Berlin: Springer, 1989, 301–311.
26. WILIE — Ein wissensbasiertes Literaturerfassungssystem. (JRJS, U.Brach, W. Wahlster, W. Woll) In: B. Endres-Niggemeyer, J. Krause (Hrsg.): *Sprachverarbeitung in Information und Dokumentation,* Berlin: Springer, 1985, 101–112.

D) SONSTIGE WISSENSCHAFTLICHE PUBLIKATIONEN

1. Zum Einfluß des Computers auf die Raumvorstellung – eine differenzielle Analyse bei Studierenden von Computerwissenschaften. (C. Quaiser-Pohl, W. Lehmann, K. Jordan, JRJS) In: Abstractband zur 6. Arbeitstagung der Fachgruppe für Differentielle Psychologie, Persönlichkeitspsychologie und Psychologische Diagnostik der Deutschen Gesellschaft für Psychologie, 13. - 14. September 2001 Universität Leipzig, Leipzig: Leipziger Univ.-Verl., 2001 (ISBN 3-935693-1), 111–112.
2. Prolegomena zum Diplomstudiengang 'Computervisualistik'. In: Der Studiengang Computervisualistik, Ankündigung, Materialien und Ordnungen, Fakultät für Informatik, Magdeburg. Otto-von-Guericke-Univ., 1996.
3. Arbeitsbericht für den Zeitraum 1991—1993: VITRA, SFB 314. (G. Herzog, JRJS, P. Wazinski) VITRA-Memo 58, SFB 314, VITRA, Univ. d. Saarlandes, Saarbrücken, 1993.
4. Zum Nutzen antizipierter Bildvorstellungen bei der sprachlichen Szenenbeschreibung. VITRA-Memo 49, SFB 314, VITRA, Universität des Saarlandes, Saarbrücken, 1991.

5. MEGA-ACT: Eine Studie über explizite Kontrollstrukturen und Pseudoparallelität in Aktor-Systemen mit einer Beispielarchitektur in FRL. KI-Lab-Memo 10, FB Informatik, Universität des Saarlandes, Saarbrücken, 1986.
6. WILIE — Ein wissensbasiertes Literaturerfassungssystem. In: *Erstes Saarbrücker Computer Forum,* Saarbrücken: IHK und die Universität des Saarlandes, 1984, 34–43. F: In Vorbereitung

E) BEITRÄGE ZU WÖRTERBÜCHERN

1. Stichpunkt „hybrid“ in *Wörterbuch der Kognitionswissenschaft*. G Strube (Ed.) Stuttgart: Klett-Cotta, 1996.